Universidad Nacional Autónoma de México

Facultad de Ciencias

Compresión, Ergodicidad y el Teorema de Shannon-McMillan-Breiman

T E S I S

QUE PARA OBTENER EL TÍTULO DE:
LIC. EN MATEMÁTICAS

PRESENTA:

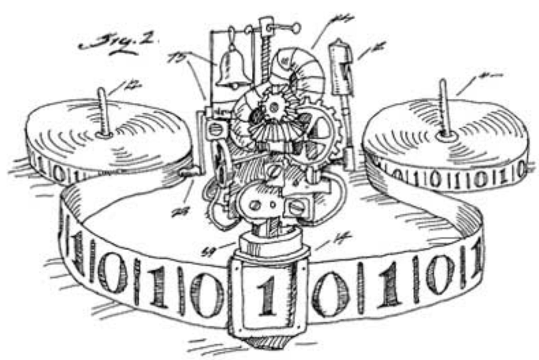
Ricardo Yadel Murillo Pérez

DIRECTOR DE TESIS:
Dr. Carlos Alfonso Cabrera Ocañas

México D.F. Ciudad Universitaria 2016

Murillo Pérez Ricardo Yadel

Compresión, Ergodicidad y el Teorema de Shannon-McMillan-Breiman



A mis padres
A María Elena Murillo

Índice general

Agradecimientos

El trabajo aquí presente es fruto del esfuerzo de muchas personas que con su apoyo contribuyeron a su desarrollo y conclusión. Sin ellos, este trabajo no hubiera sido posible. Los aciertos que pueda haber en él son el resultado de sus comentarios y aportaciones. Sin embargo, los errores que pudiera haber son sólo míos.

En primer lugar, agradezco a mis padres, Alma Rosa Pérez Pérez y Rubén Murillo Loza por su incondicional apoyo. Así también agradezco de manera particular a Brenda Murillo y a Lizbeth Ramírez por sus consejos y apoyo constante. Asimismo agradezco a Rubén Pavel Murillo y Beatríz Rodríguez por sus comentarios y sugerencias. Agradezco a los compañeros y amigos que me dieron su apoyo y palabras de aliento.

Agradezco también a mi institución, a la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM) y a todos los excelentes profesores con quienes tuve la oportunidad de asistir a sus cursos, a quienes les debo mi formación personal y académica.

Introducción

La gran mayoría de las personas actualmente está familiarizada con los *archivos comprimidos*. Existen programas de computadora que reducen el espacio que ocupan los archivos para facilitar su descarga, o bien, simplemente para transportarlos a otra computadora de una manera más eficiente. Una vez que se ha descargado o transportado el programa comprimido, el mismo programa que nos ayudó a comprimir permite a su vez realizar la operación inversa que es la descompresión para permitir la manipulación del archivo.

El proceso de compresión y descompresión que se realiza en una computadora tiene un fundamento matemático interesante. Los archivos en una computadora se pueden ver como sucesiones finitas de símbolos [21]. Para comprimir un archivo representamos cada símbolo del archivo por una secuencia de ceros y unos. Este proceso se conoce como *codificación del archivo*. Al reemplazar cada secuencia de ceros y unos en la sucesión original observamos que el número de ceros y unos en la nueva sucesión depende de la manera en que se realizó la codificación. Dado que cada símbolo ocupa un espacio en la memoria de la computadora, una mejor codificación será aquella de la que resulte una sucesión más corta. El *problema de la compresión* consiste en encontrar, para una cadena de símbolos dada, aquella codificación que arroje sucesiones de ceros y unos lo más cortas posibles, en cuyo caso podremos hablar de codificaciones óptimas. En este sentido, el problema de la compresión es un *problema de optimización*.

La idea central que subyace para alcanzar una codificación óptima consiste primero en calcular la frecuencia con la que ocurre cada símbolo en la sucesión. Posteriormente, en asignar secuencias más cortas a los símbolos más frecuentes, y en asignar las secuencias más largas a los símbolos menos frecuentes.

La manera en que se realiza la compresión de archivos nos lleva de manera natural a hacernos dos preguntas fundamentales. La primera pregunta que surge es si existe un "límite" a la compresión, es decir, si dada una sucesión de símbolos se puede establecer una cota a la longitud de una sucesión codificada tal que ninguna codificación posible podría tener una longitud menor. La existencia de dicha cota nos podría sugerir la existencia de *códigos óptimos*. La segunda pregunta tiene que ver con la manera "práctica" de obtener los mejores códigos posibles, es decir, en desarrollar algoritmos que nos ayuden a comprimir archivos. Para dar una respuesta a la primera pregunta recurriremos a dos conceptos clave, como son el concepto de *entropía* y el concepto de *ergodicidad*. La respuesta a la segunda pregunta se abordará hasta el capítulo 5 al presentar el *algoritmo de*

Huffman para comprimir datos.

El concepto de ergodicidad se desarrollará en los primeros tres capítulos de este trabajo. En los primeros tres capítulos, se muestran los principales resultados de la Teoría Ergódica sobre los sistemas dinámicos en general. En el primer capítulo se define el concepto de *transformación preservadora de la medida* y se muestran algunos de los ejemplos más ilustrativos de este tipo de transformaciones. Posteriormente en el mismo capítulo se aborda el *problema de la existencia de las medidas invariantes*. Una vez que hemos encontrado algunas transformaciones que preservan la medida, cabe preguntarse si siempre existen medidas para transformaciones dadas tales que las transformaciones sean transformaciones preservadoras. La solución al Problema de la existencia de medidas invariantes se condensa en el *Teorema de Kyrilov-Bogolubov*. El Teorema de Kyrilov-Bogolubov asegura que siempre se puede encontrar dicha medida siempre y cuando el espacio medible sea un espacio métrico compacto.

En el segundo capítulo se prueban los teoremas centrales de la Teoría Ergódica. Por una parte, se prueba el *Teorema de la Recurrencia de Poincaré* y, por otra parte, se prueba el *Teorema Ergódico de Birkhoff*. El Teorema de la Recurrencia de Poincaré afirma que la órbita de cada punto en un conjunto medible regresa un número infinito de veces al conjunto medible. Sin embargo, sería deseable saber también cuanto tiempo permanecen las órbitas en el conjunto. Dicha respuesta nos la ofrece el Teorema Ergódico de Birkhoff al sostener que la frecuencia con la que cae la órbita de un punto en un conjunto medible está dado por la *esperanza condicional* de la *función característica* del conjunto en cuestión con respecto a la *σ-álgebra* de los *conjuntos invariantes*. En este sentido, el Teorema Ergódico de Birkhoff completa la respuesta que nos da el Teorema de la Recurrencia de Poincaré.

En el capítulo siguiente se desarrolla de manera exhaustiva el concepto de ergodicidad. La noción de ergodicidad nos describe un comportamiento especial de las órbitas de los puntos en el espacio, a saber, que cada órbita cae en *casi todos* los conjuntos medibles. Dicho comportamiento trae consecuencias importantes para el Teorema Ergódico de Birkhoff, pues nos facilita los cálculos. Así también, se desarrollan algunos ejemplos y se verifican las condiciones de ergodicidad de algunas de las transformaciones que preservan la medida más importantes. Finalmente se termina el capítulo 3 con el estudio de la *ergodicidad única*. El Teorema de Kyrilov-Bogolubov asegura la existencia de medidas invariantes. Sin embargo, en ocasiones la medida invariante es única y en este caso decimos que la transformación es únicamente ergódica. La ergodicidad única fortalece a su vez el Teorema Ergódico de Birkhoff al asegurar no sólo la convergencia puntual, sino que se puede asegurar que la convergencia es uniforme.

El concepto de entropía se desarrolla en el penúltimo capítulo. En el capítulo 4 primero se define la entropía en relación con una partición medible, posteriormente la entropía con respecto a una partición y a una transformación medible y finalmente la entropía de una transformación y se prueban sus principales propiedades. Posteriormente se muestran algunos ejemplos y finalmente se prueba, con ayuda del Teorema Ergódico de Birkhoff, el *Teorema de*

Shannon-McMillan-Breiman.

Por último, en el capítulo 5, se prueba, a partir del Teorema de Shannon-McMillan-Breiman, la llamada Propiedad de la Equipartición Asintótica (AEP). La AEP afirma que podemos encontrar conjuntos de alta probabilidad y conjuntos de baja probabilidad. Los conjuntos de baja probabilidad pueden ser ignorados para trabajar únicamente con los de alta probabilidad. La AEP tiene importantes consecuencias en términos de aplicaciones. En el caso de la compresión de datos, nos permite probar que la tasa de compresión máxima coincide con la entropía de la transformación, que en este caso se trata de la *transformación corrimiento* con respecto a la *medida de Bernoulli*. Esto quiere decir que bajo una codificación óptima, es decir, la codificación más corta que se podría hacer de una sucesión de símbolos, cada símbolo tiene en promedio una longitud cercana a la entropía. Para terminar el capítulo, se muestra un algoritmo que nos permite encontrar un código óptimo, a saber, el *algoritmo de Huffman*, y se ilustra su funcionamiento mediante un programa de computadora. El programa que se presenta fue realizado en lenguaje C utilizando *estructuras dinámicas de datos*.

A lo largo de este trabajo se hacen constantes referencias a resultados y conceptos que provienen principalmente de la Teoría de la Medida. Para agilizar la lectura simplemente se enuncian antes de la prueba en la que se ocupan. Las pruebas se omiten porque se trata de resultados que requieren de una construcción elaborada y se deben examinar a detalle. El trabajo del esclarecimiento de dichos conceptos es una labor loable pero excede los alcances de este trabajo. Los trabajos [11], [3] y [12] constituyen excelentes referencias para consultar los resultados de la Teoría de la Medida.

Capítulo 1

Transformaciones Preservadoras de la Medida

En este capítulo se introduce la noción básica de la *Teoría Ergódica*, a saber, la de *transformación que preserva una medida* y se muestran algunos de los ejemplos más ilustrativos. Posteriormente se aborda el *Problema de la existencia de medidas invariantes.* El problema consiste en determinar si para toda transformación T de un espacio X en sí mismo, existe una medida μ tal que es *preservada* por T. Posteriormente se muestra que bajo las hipótesis de que el espacio X sea un espacio métrico compacto y la transformación T sea una transformación continua, el problema siempre tiene solución.

1.1. Medidas Invariantes

Los espacios que se consideran a lo largo de este trabajo tienen estructura de *espacios de medida.* Así pues, consideremos un conjunto no vacío, una σ-álgebra S definida sobre los subconjuntos de X y una medida μ. Nos interesa trabajar con transformaciones que preserven la estructura del espacio, por ello, trabajaremos con las transformaciones medibles. De manera formal, dado un espacio de medida (X, S, μ) y una transformación $T : X \to X$, decimos que T es una ***transformación medible*** si la imagen inversa de un conjunto en S es de nuevo un conjunto en S, es decir:

$$T^{-1}(B) = \{x \in X : T(x) \in B\} \in S \qquad \forall B \in S. \tag{1.1}$$

A continuación, introducimos la noción de *transformación preservadora de la medida*, o bien, *medida invariante* respecto a una transformación medible. Una transformación T preserva una medida μ si al tomar la imagen inversa de un conjunto medible, mide lo mismo que el conjunto medible.

Definición 1. *Sea (X, S, μ) un espacio de medida y sea $T : X \to X$ una transformación medible. Decimos que μ es T-**invariante**, o bien, que T **preserva***

la medida *μ si*

$$\mu(T^{-1}(B)) = \mu(B) \qquad \forall B \in S.$$

Cuando la transformación T es invertible y su inversa T^{-1} es también una transformación medible, la condición para que T preserve la medida μ es equivalente a la condición:

$$\mu(T(B)) = \mu(B) \qquad \forall B \in S.$$

1.2. Ejemplos de Transformaciones que Preservan la Medida

En esta sección estudiaremos algunos de los principales ejemplos de transformaciones que preservan la medida. Para probar que las transformaciones son preservadoras ocuparemos un resultado que afirma que para verificar si una transformación preserva una medida, no hace falta verificar la condición $\mu(T^{-1}(B)) = \mu(B)$ para cada conjunto B de la σ-álgebra, sino sólo para los conjuntos del *álgebra* que la genera. Recordamos que un álgebra de conjuntos consiste en subconjuntos cerrados bajo diferencias y bajo uniones.

Definición 2. *Sea X un conjunto no vacío. Decimos que $\mathcal{A} \subseteq \mathcal{P}(X)$ es un* ***álgebra de conjuntos*** *si se cumple lo siguiente:*

1. *$X \in \mathcal{A}$.*
2. *Si $E, F \in \mathcal{A}$, entonces $E \setminus F \in \mathcal{A}$.*
3. *Si $E, F \in \mathcal{A}$, entonces $E \cup F \in \mathcal{A}$.*

El resultado central que ocuparemos para probar el teorema auxiliar es el *Lema de las Clases Monótonas*, el cual se enuncia en términos de *π-sistemas* y *λ-sistemas*. Un π-sistema y un λ-sistema se definen de la siguiente manera.

Definición 3. *Sea X un conjunto no vacío.*

- *Decimos que $C \subseteq \mathcal{P}(X)$ es un* ***π-sistema*** *si dados $E, F \in C$, entonces se tiene que $E \cap F \in C$.*
- *Decimos que $L \subseteq \mathcal{P}(X)$ es un* ***λ-sistema*** *si se cumplen las siguientes condiciones:*
    1. *$X \in L$.*
    2. *Si se tienen $E, F \in L$ tales que $F \subseteq E$, entonces $E \setminus F \in L$.*
    3. *Si $(E_n)_{n=1}^{\infty} \subseteq L$ es una sucesión creciente[1], entonces $\bigcup_{n=1}^{\infty} E_n \in L$.*

[1] Es decir, $E_n \subseteq E_{n+1} \forall n \in \mathbb{N}$

Enunciamos sin prueba el Lema de las Clases Monótonas y a continuación el teorema que ocuparemos en los ejemplos para probar que las transformaciones preservan la medida. Dado un conjunto $C \subseteq \mathcal{P}(X)$, denotamos la σ-álgebra generada por C, es decir, la σ-álgebra mas pequeña que contiene a C, como $S(C)$.

Teorema 1 (Lema de las Clases Monótonas)**.** *Sea X un conjunto no vacío y $C \subseteq \mathcal{P}(X)$ un π-sistema y $L \subseteq \mathcal{P}(X)$ un λ-sistema tal que $C \subseteq L$. Entonces se tiene que $S(C) \subseteq L$.*

Teorema 2. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$ y sea $T \colon X \to X$ una transformación medible. Si existe un álgebra $\mathcal{A}$ tal que $S(\mathcal{A}) = S$ y que cumple,*

$$\mu(T^{-1}(A)) = \mu(A) \qquad \forall A \in \mathcal{A}.$$

entonces T preserva la medida μ, es decir,

$$\mu(T^{-1}(B)) = \mu(B) \qquad \forall B \in S.$$

Demostración. Consideremos el siguiente conjunto:

$$L = \{E \in S : \mu(T^{-1}(E)) = \mu(E)\}.$$

Por hipótesis se tiene que $\mathcal{A} \subseteq L \subseteq S$ y que $S(\mathcal{A}) = S$. Por el Lema de las Clases Monótonas basta verificar:

- $\mathcal{A}$ es un π-sistema.
- L es un λ-sistema.

Veamos que $\mathcal{A}$ es un π-sistema. Sean $E, F \in \mathcal{A}$. Por definición de álgebra, se sabe que es cerrada bajo uniones y diferencias. Así pues, para ver que $E \cap F \in \mathcal{A}$, basta probar la siguiente afirmación:
Afirmación: $E \cap F = (E \cup F) \setminus ((E \setminus S) \cup (F \setminus E))$.

$$\begin{aligned}
(E \cup F) \setminus ((E \setminus F) \cup (F \setminus E)) &= (E \cup F) \setminus ((E \cap F^c) \cup (F \cap E^c)) \\
&= (E \cup F) \cap ((E \cap F^c) \cup (F \cap E^c))^c \\
&= (E \cup F) \cap ((E \cap F^c)^c \cap (F \cap E^c)^c) \\
&= (E \cup F) \cap (E^c \cup F) \cap (F^c \cup E) \\
&= (F \cup (E^c \cap E)) \cap (F^c \cup E) \\
&= (F \cap F^c) \cup (E \cap F) \\
&= E \cap F.
\end{aligned}$$

Veamos que L es un λ-sistema. Verificamos la primera condición. Como $T^{-1}(X) = X$, entonces se tiene que $\mu(T^{-1}(X)) = \mu(X) = 1$. Por lo tanto, $X \in L$.

Verificamos la segunda condición. Consideramos $E, F \in L$ tales que $F \subseteq E$ y veamos que $E \setminus F \in L$.

$$\begin{aligned}\mu(T^{-1}(E \setminus F)) &= \mu(T^{-1}(E) \setminus T^{-1}(F)) \\ &= \mu(T^{-1}(E)) - \mu(T^{-1}(F)) \\ &= \mu(E) - \mu(F) \\ &= \mu(E \setminus F).\end{aligned}$$

Luego pues, $E \setminus F \in L$. Verificamos la tercera condición. Sea $(E_n)_{n=1}^{\infty} \subseteq L$ una sucesión creciente, entonces,

$$\begin{aligned}\mu\left(T^{-1}\left(\bigcup_{n=1}^{\infty} E_n\right)\right) &= \mu\left(\bigcup_{n=1}^{\infty} T^{-1}(E_n)\right) \\ &= \lim_{n\to\infty} \mu(T^{-1}(E_n)) \\ &= \lim_{n\to\infty} \mu(E_n) \\ &= \mu\left(\bigcup_{n=1}^{\infty} E_n\right).\end{aligned}$$

Concluimos que $\bigcup_{n=1}^{\infty} E_n \in L$. Así pues, L es un λ-sistema. Por el Lema de las Clases Monótonas, tenemos que,

$$S = S(\mathcal{A}) \subseteq L \subseteq S.$$

Por lo tanto, $L = S$. Así pues T preserva la medida μ. $\square$

1.2.1. Transformación Expansión Binaria

Consideremos el espacio de medida $([0,1], S, \lambda)$, donde S es la σ-álgebra de Borel y λ es la medida de Lebesgue. Sea $m \in \mathbb{N}$, la transformación $T : X \to X$ definida como sigue recibe el nombre de ***transformación expansiva lineal a pedazos***:

$$T(x) = mx \mod 1 = \begin{cases} mx & \text{si} \quad 0 \le x < \frac{1}{m} \\ mx - 1 & \text{si} \quad \frac{1}{m} \le x < \frac{2}{m} \\ mx - 2 & \text{si} \quad \frac{2}{m} \le x < \frac{3}{m} \\ \vdots & \vdots \quad \vdots \\ mx - m & \text{si} \quad \frac{m-1}{m} \le x < 1. \end{cases}$$

Observamos que para cada $m \in \mathbb{N}$ se define una transformación expansiva lineal a pedazos diferente. Para entender como se comportan las transformaciones expansivas lineales a pedazos, utilizaremos el caso particular cuando $m = 2$. Este caso recibe el nombre de ***transformación expansión binaria***.

$$T(x) = 2x \mod 1 = \begin{cases} 2x & \text{si} \quad 0 \le x < \frac{1}{2} \\ 2x - 1 & \text{si} \quad \frac{1}{2} \le x < 1. \end{cases} \tag{1.2}$$

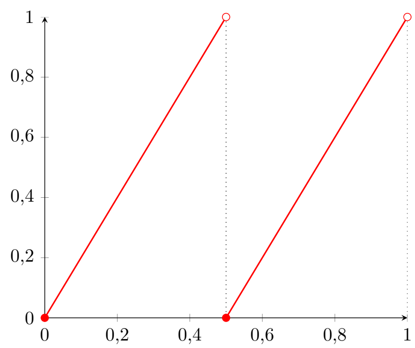


Figura 1.1: $T(x) = 2x \mod 1$.

A continuación probaremos que la transformación expansión binaria preserva la medida de Lebesgue.

Proposición 1. *La transformación $T : [0,1) \to [0,1)$ definida como en (1.2) preserva la medida de Lebesgue.*

Demostración. Para ver que la transformación $T(x) = 2x \mod 1$ preserva la medida de Lebesgue, se debe verificar que $\lambda(T^{-1}(A)) = \lambda(A)$ para todo Boreliano $A \subseteq [0,1)$. Ahora bien, por el Teorema 2, se sabe que basta probar que T preserva la medida para los generadores de los Borelianos, que son los intervalos de la forma $[a,b)$. Asimismo, como $[a,b) = [0,b) \setminus [0,a)$, basta probar que ocurre lo siguiente:

$$\lambda(T^{-1}([0,a))) = \lambda([0,a)) \qquad \forall a \in (0,1).$$

Observamos que $T^{-1}([0,a)) = [0, \frac{a}{2}) \cup [\frac{1}{2}, \frac{a+1}{2})$. Como la unión es disjunta, al tomar la medida se tiene lo siguiente:

$$\begin{aligned}\lambda(T^{-1}([0,a))) &= \lambda\left(\left[0, \frac{a}{2}\right)\right) + \lambda\left(\left[\frac{1}{2}, \frac{a+1}{2}\right)\right) \\ &= \frac{a}{2} + \frac{a}{2} \\ &= a \\ &= \lambda([0,a)).\end{aligned}$$

Por lo tanto, T preserva la medida de Lebesgue. $\square$

La función $T(x) = 2x \mod 1$ no sólo es el caso más simple que ilustra la manera en que se comportan las transformaciones lineales a pedazos, sino que tiene aplicaciones importantes en la Teoría de Números. En el siguiente capítulo se mostrará la relación de la transformación T con las expansiones binarias de un número real.

1.2.2. Rotaciones en el Círculo

Consideremos el conjunto $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ y observamos que $\mathbb{S}^1$ se puede identificar con el conjunto $\Pi = [0,1]/\{0,1\}$ mediante el *isomorfísmo* $h : \Pi \to \mathbb{S}^1$ definido como sigue:

$$h(\theta) = e^{2\pi i\theta}.$$

Ahora bien, para $w \in \mathbb{S}^1$ se define la función $R_w : \mathbb{S}^1 \to \mathbb{S}^1$ dada por:

$$R_w(z) = wz.$$

La función R_w se conoce como ***rotación en el círculo***. Asimismo, definimos la función $T_\alpha : \Pi \to \Pi$ de la siguiente manera:

$$T_\alpha(x) = x + \alpha \mod 1.$$

A la función T_α se le conoce como ***traslación del intervalo***. Veamos que el siguiente diagrama conmuta.

$$\begin{tikzcd}
\Pi \arrow[r, "T_\alpha"] \arrow[d, "h"'] & \Pi \arrow[d, "h"] \\
S^1 \arrow[r, "R_w"] & S^1
\end{tikzcd}$$

Sea α un número real contenido en el intervalo $[0,1]$. Consideramos $w = e^{2\pi i\alpha}$. Para cada $\theta \in \Pi$ se tiene que,

$$\begin{aligned}
R_w \circ h(\theta) &= R_w(e^{2\pi i\theta}) \\
&= e^{2\pi i\alpha}e^{2\pi i\theta} \\
&= e^{2\pi i(\alpha+\theta)} \\
&= h(\alpha + \theta) \\
&= h \circ T_\alpha(\theta).
\end{aligned} \tag{1.3}$$

Primero probaremos que la transformación T_α preserva la medida m en Π inducida por la medida de Lebesgue en $[0,1]$. Posteriormente consideraremos la medida de Lebesgue λ inducida en el espacio $\mathbb{S}^1$ y probaremos que para cada $w \in \mathbb{C}$ la rotación R_w preserva la medida λ.

Proposición 2. *Sea $\alpha \in (0,1)$. La traslación T_α preserva la medida m en Π inducida por la medida de Lebesgue en $[0,1]$.*

Demostración. Sea $B \subseteq \Pi$ un conjunto medible. Por el Teorema 2, sabemos que basta probar que preserva la medida para los generadores de los Borelianos. Así pues, basta probar que ocurre:

$$m(T_\alpha^{-1}([0,a))) = m([0,a)) \qquad \forall a \in (0,1).$$

Tenemos dos casos:

Caso 1: $a < \alpha$.

$$T_\alpha^{-1}([0,a)) = [1-\alpha, a-\alpha+1).$$

Esto implica que $m(T_\alpha^{-1}([0,a))) = m([1-\alpha, a-\alpha+1)) = a = m([0,a))$.

Caso 2: $a > \alpha$.

$$T_\alpha^{-1}([0,a)) = [0, a-\alpha) \cup [1-\alpha, 1).$$

Luego pues, $m(T_\alpha^{-1}([0,a))) = m([0,a-\alpha) \cup [1-\alpha,1)) = a = m([0,a))$. Por lo tanto, T_α preserva la medida m. $\square$

Ahora bien, a partir de la medida m que es la medida en Π inducida por la medida de Lebesgue en $[0,1]$, se puede inducir una medida λ en el espacio $\mathbb{S}^1$ como sigue,

$$\lambda(h(B)) = m(B) \qquad \forall B \subseteq \Pi \text{ conjunto medible.}$$

Llamamos a la medida λ la medida de Lebesgue en $\mathbb{S}^1$. La siguiente proposición prueba que las rotaciones en el círculo preservan la medida de Lebesgue en $\mathbb{S}^1$.

Proposición 3. *Sea $w \in \mathbb{C}$. La rotación R_w preserva la medida de Lebesgue λ en $\mathbb{S}^1$.*

Demostración. Por (1.3) se sabe que,

$$R_w(h(B')) = h(T_\alpha(B')) \qquad \forall B' \subseteq \Pi \text{ conjunto medible.}$$

Así pues, sea $B \subseteq \Pi$ un conjunto medible. Tenemos lo siguiente,

$$\begin{aligned} \lambda(R_w(h(B))) &= \lambda(h(T_\alpha(B))) \\ &= m(T_\alpha(B)). \end{aligned}$$

Como T_α es invertible, por la Proposición 2 se tiene que $m(T_\alpha(B)) = m(B) = \lambda(h(B))$. Por lo tanto la transformación R_w preserva la medida λ. $\square$

1.2.3. Transformación de Gauss

Consideremos la transformación $T_G : [0,1) \to [0,1)$ definida de la siguiente manera:

$$T_G(x) = \begin{cases} \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor & \text{si} \quad x \neq 0 \\ 0 & \text{si} \quad x = 0, \end{cases}$$

donde $\left\lfloor \frac{1}{x} \right\rfloor$ denota la “parte entera” de $\frac{1}{x}$. La transformación que se acaba de definir recibe el nombre de ***transformación de Gauss***.

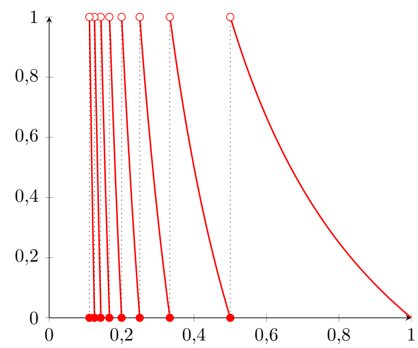


Figura 1.2: Transformación de Gauss.

La transformación de Gauss es importante al considerar la expansión de un número irracional, pues sus iteraciones generan la expansión en *fracciones continuas* para cada punto en $(0,1)$. Dado $x \in (0,1) \cap \mathbb{I}$, definimos la función:

$$n_j(x) = \left\lfloor \frac{1}{T_G^{j-1}(x)} \right\rfloor \qquad \forall j \in \mathbb{N}.$$

Los n_j son los *cocientes parciales* de la expansión en fracciones continuas de x. Observamos lo siguiente:

$$\begin{aligned} T_G^j(x) &= T_G(T_G^{j-1}(x)) \\ &= \frac{1}{T_G^{j-1}(x)} - n_j(x). \end{aligned}$$

Esto implica que

$$T_G^{j-1}(x) = \frac{1}{T_G^j(x) + n_j(x)} \qquad \forall j \in \mathbb{N},$$

es decir,

$$\begin{aligned} x &= \frac{1}{n_1(x) + T_G(x)} \\ &= \frac{1}{n_1(x) + \cfrac{1}{n_2(x) + T_G^2(x)}} \\ &= \frac{1}{n_1(x) + \cfrac{1}{n_2(x) + \cfrac{1}{n_3(x) + T_G^3(x)}}}. \end{aligned}$$

La transformación de Gauss no preserva la medida de Lebesgue, pero sí preserva la medida que se definirá a continuación y que recibe el nombre de ***medida de Gauss***. Si consideramos el espacio $X = [0,1)$ y la σ-álgebra de Borel S, podemos definir la función $\mu_G : S \to \mathbb{R}$ como sigue:

$$\mu_G(A) = \frac{1}{\log 2} \int_A \frac{\mathrm{d}x}{1+x} \qquad \forall A \in S. \tag{1.4}$$

Proposición 4. *La función $\mu_G : S \to \mathbb{R}$ definida en (1.4) es una medida.*

Demostración. Para ver que la función $\mu_G : S \to \mathbb{R}$ es una medida, se debe verificar que se cumplan las siguientes propiedades:

- $\mu(\emptyset) = 0$.
- $\mu(E) \geq 0$, para todo $E \in S$.
- Si $(E_n)_{n\in\mathbb{N}} \subseteq S$ es una sucesión de elementos disjuntos, entonces

$$\mu_G\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu_G(E_n).$$

Veamos que $\mu_G(\emptyset) = 0$.

$$\begin{aligned}\mu(\emptyset) &= \frac{1}{\log 2} \int_\emptyset \frac{\mathrm{d}x}{1+x} \\ &= \frac{1}{\log 2}(0) = 0.\end{aligned}$$

Ahora, veamos que $\mu_G(E) \geq 0$, para todo $E \in S$. Sea $E \in S$. Como la función logaritmo es una función creciente y positiva, se tiene lo siguiente:

$$\begin{aligned}\mu_G(E) &= \frac{1}{\log 2} \int_E \frac{\mathrm{d}x}{1+x} \\ &= \frac{1}{\log 2}(\log(1+x))\,|_E \\ &\geq 0.\end{aligned}$$

Finalmente, supongamos que $(E_n)_{n\in\mathbb{N}} \subseteq S$ es una sucesión de elementos disjuntos. Se tiene que,

$$\begin{aligned}\mu_G\left(\bigcup_{n=1}^{\infty} E_n\right) &= \frac{1}{\log 2} \int_{\bigcup_{n=1}^{\infty} E_n} \frac{\mathrm{d}x}{1+x} \\ &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{E_n} \frac{\mathrm{d}x}{1+x} \\ &= \sum_{n=1}^{\infty} \mu_G(E_n).\end{aligned}$$

Concluimos que μ_G es una medida. $\square$

A continuación se mostrará que la transformación de Gauss preserva la medida de Gauss.

Teorema 3. *La transformación de Gauss preserva la medida de Gauss.*

Demostración. Veamos que la transformación de Gauss preserva la medida de Gauss, es decir, que $\mu_G(T_G^{-1}(A)) = \mu(A)$ para todo Boreliano $A \subseteq [0,1)$. Por el Teorema 2 basta probar que ocurre:

$$\mu_G(T_G^{-1}([0,s))) = \mu([0,s)) \qquad \forall s \in (0,1).$$

Afirmación: $T_G^{-1}([0,s)) = \bigcup_{n=1}^{\infty}[\frac{1}{n+s}, \frac{1}{n})$.

$$\begin{aligned} x \in \bigcup_{n=1}^{\infty}\left[\frac{1}{n+s}, \frac{1}{n}\right) &\Leftrightarrow \frac{1}{n+s} \leq x < \frac{1}{n} \qquad \text{para alguna } n \in \mathbb{N} \\ &\Leftrightarrow T_G\left(\frac{1}{n}\right) \leq T_G(x) \leq T_G\left(\frac{1}{n+s}\right) \qquad \text{para alguna } n \in \mathbb{N} \\ &\Leftrightarrow n - n \leq T_G(x) \leq n + s - n \qquad \text{para alguna } n \in \mathbb{N} \\ &\Leftrightarrow 0 \leq T_G(x) \leq s. \end{aligned}$$

Como $\bigcup_{n=1}^{\infty}[\frac{1}{n+s}, \frac{1}{n})$ es una unión de elementos disjuntos, por la σ-aditividad de μ_G, se tiene lo siguiente:

$$\begin{aligned} \mu_G(T_G^{-1}([0,s))) &= \sum_{n=1}^{\infty} \mu_G\left[\frac{1}{n+s}, \frac{1}{n}\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{\log 2} \int_{\frac{1}{n+s}}^{\frac{1}{n}} \frac{\mathrm{d}x}{1+x} \\ &= \sum_{n=1}^{\infty} \frac{1}{\log 2} \log(1+x) \Big|_{\frac{1}{n+b}}^{\frac{1}{n}} \\ &= \sum_{n=1}^{\infty} \frac{1}{\log 2} \left(\log\left(1 + \frac{1}{n}\right) - \log\left(1 - \frac{1}{n+s}\right)\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{\log 2} \left(\log \frac{n+1}{n+s+1} - \log \frac{n}{n+s}\right) \\ &= \frac{1}{\log 2}\left(\sum_{n=1}^{\infty} \log \frac{n+1}{n+s+1} - \sum_{n=1}^{\infty} \log \frac{n}{n+s}\right) \\ &= \frac{1}{\log 2}\left(\sum_{n=2}^{\infty} \log \frac{n}{n+s} - \sum_{n=2}^{\infty} \log \frac{n}{n+s} - \log \frac{1}{1+s}\right) \\ &= \frac{1}{\log 2}\left(-\log \frac{1}{1+s}\right) \\ &= \frac{1}{\log 2} \int_0^s \frac{\mathrm{d}x}{1+x} \end{aligned}$$

$$= \mu_G([0, s)).$$

□

1.2.4. Endomorfismos en el Toro

Comenzaremos definiendo lo que se entenderá por un *toro n'ésimo* y después caracterizaremos los endomorfismos. Un toro n-ésimo es un conjunto que consiste de las clases de equivalencia inducidas por la siguiente relación $\sim$

$$x \sim y \Leftrightarrow x - y \in \mathbb{Z}^n \qquad \forall x \in \mathbb{R}^n.$$

Denotamos las clases de equivalencia como:

$$[x] = \{y \in \mathbb{R}^n : x \sim y\}.$$

Finalmente, el ***toro n-ésimo*** se define como un conjunto de clases de equivalencia,

$$\begin{aligned} \mathbb{T}^n &= \mathbb{R}^n / \mathbb{Z}^n \\ &= \{[x] : x \in \mathbb{R}^n\}. \end{aligned}$$

Antes de definir las transformaciones que nos interesan, veremos algunas propiedades importantes del conjunto de matrices invertibles.

$$G(n, \mathbb{Z}) = \{A \in M_{n \times n}(\mathbb{Z}) : \det(A) \neq 0\}.$$

Notamos que $A(\mathbb{Z}^n) \subseteq \mathbb{Z}^n$ para toda $A \in G(n, \mathbb{Z})$. Sabemos que cada $A \in G(n, \mathbb{Z})$ define una transformación lineal sobre $\mathbb{R}^n$. Si $y - x \in \mathbb{Z}^n$, como $A(\mathbb{Z}^n) \subseteq \mathbb{Z}^n$, entonces $Ay - Ax \in \mathbb{Z}^n$. Es decir, si $y - x \in \mathbb{Z}^n$, entonces $Ay - Ax \in \mathbb{Z}^n$. Como $y - x \in \mathbb{Z}^n$ implica que $Ay - Ax \in \mathbb{Z}^n$, entonces,

$$\begin{aligned} \text{si } y \in [x] &\Rightarrow y - x \in \mathbb{Z}^n \\ &\Rightarrow Ay - Ax \in \mathbb{Z}^n \\ &\Rightarrow Ay \in [Ax]. \end{aligned}$$

Definimos un "endomorfismo en el toro" como sigue.

Definición 4. *Sea $A \in G(n, \mathbb{Z})$. Definimos un* ***endomorfismo en el toro*** *$T_A : \mathbb{T}^n \to \mathbb{T}^n$ como:*

$$T_A([x]) = [Ax] \qquad \forall [x] \in \mathbb{T}^n.$$

La siguiente proposición nos da una caracterización de la inversa de un endomorfismo en el toro T_A cuando esta exista.

Proposición 5. *Sea $A \in G(n, \mathbb{Z})$ y $T_A : \mathbb{T}^n \to \mathbb{T}^n$ un endomorfismo en el toro invertible, entonces*

$$(T_A)^{-1} = T_{A^{-1}}.$$

Demostración.

$$\begin{aligned}(T_{A^{-1}} \circ T_A)([x]) &= T_{A^{-1}}(T_A(([x]))) \\ &= T_{A^{-1}}([Ax]) \\ &= [A^{-1}Ax] = [x].\end{aligned}$$

Análogamente $(T_A \circ T_{A^{-1}})([x]) = [x]$. □

En general T_A no necesariamente es invertible aún cuando $A \in G(n, \mathbb{Z})$ sea invertible.

Ejemplo 1 (Endomorfismo en el toro que no es invertible)**.** Consideremos:

$$A = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}.$$

Observamos que $\det A = 2 \neq 0$. Esto implica que $A \in G(2, \mathbb{Z})$, es decir, A es invertible. Sin embargo, T_A no es invertible, pues,

$$\begin{aligned}A^{-1} &= \frac{1}{\det A}\begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & -1 \\ -1 & \frac{3}{2} \end{pmatrix}.\end{aligned}$$

Es decir, $A^{-1} \notin M_{2\times 2}(\mathbb{Z})$ y entonces $A^{-1} \notin G(n, \mathbb{Z})$.

Ahora, consideremos el conjunto de las matrices cuadradas con coeficientes en los enteros y cuyo determinante vale 1 o -1.

$$S(n, \mathbb{Z}) = \{A \in M_{n\times n}(\mathbb{Z}) : |\det A| = 1\}.$$

Observamos que $S(n, \mathbb{Z}) \subseteq G(n, \mathbb{Z})$. La siguiente proposición nos da condiciones necesarias y suficientes para que una transformación T_A sea invertible.

Proposición 6. *T_A es invertible si y sólo si $A \in S(n, \mathbb{Z})$.*

Demostración. Supongamos que T_A es invertible, entonces existe $(T_A)^{-1}$ tal que

$$T_A \circ (T_A)^{-1} = (T_A)^{-1} \circ T_A = Id.$$

Por la Proposición 5 se tiene que,

$$(T_A)^{-1} = T_{A^{-1}}.$$

Entonces $A \in G(n, \mathbb{Z})$ y también $A^{-1} \in G(n, \mathbb{Z})$. Esto implica que,

$$|\det A^{-1}| = 1.$$

Así también,

$$|\det A| = 1.$$

Por lo tanto $A \in S(n, \mathbb{Z})$. Recíprocamente supongamos que $A \in S(n, \mathbb{Z})$, entonces $|\det A| = 1$. Esto implica que $A^{-1} \in M_{n\times n}(\mathbb{Z})$. Así pues, $T_{A^{-1}}$ es un endomorfismo en el toro y como $T_{A^{-1}} = (T_A)^{-1}$, concluimos que T_A es invertible. □

La siguiente proposición nos muestra una equivalencia mas para saber si un endomorfismo en el toro T_A es invertible.

Proposición 7. *Una condición necesaria y suficiente para que T_A sea invertible es que*

$$Ay - Ax \in \mathbb{Z}^n \Leftrightarrow y - x \in \mathbb{Z}^n.$$

Demostración. Por la Proposición 6 tenemos que,

$$\begin{aligned} T_A \text{ es invertible } &\Leftrightarrow A \in S(n, \mathbb{Z}) \\ &\Leftrightarrow |\det A| = 1 \\ &\Leftrightarrow A^{-1} \in M_{n\times n}(\mathbb{Z}) \\ &\Leftrightarrow A^{-1}(\mathbb{Z}^n) \subseteq \mathbb{Z}^n \\ &\Leftrightarrow \mathbb{Z}^n \subseteq A(\mathbb{Z}^n). \end{aligned}$$

Como $A(\mathbb{Z}^n) \subseteq \mathbb{Z}^n$, se tiene que $\mathbb{Z}^n = A(\mathbb{Z}^n)$. Pero ésto ocurre si y sólo si

$$A(y - x) \in \mathbb{Z}^n \Leftrightarrow y - x \in \mathbb{Z}^n.$$

Esto ocurre a su vez si y sólo si,

$$Ay - Ax \in \mathbb{Z}^n \Leftrightarrow y - x \in \mathbb{Z}^n.$$

□

A continuación, vamos a definir un tipo especial de endomorfismo que llamaremos los automorfismos en el toro n'ésimo.

Definición 5. *Para cada $A \in S(n, \mathbb{Z})$, el endomorfismo $T_A : \mathbb{T}^n \to \mathbb{T}^n$ recibe el nombre de* ***automorfismo en el toro****.*

Hacemos las siguientes observaciones:

1. Por la proposición anterior, se tiene que todo automorfismo en el toro es invertible.

2. La inversa de un automorfismo en el toro T_A es también un automorfismo en el toro.

Ejemplo 2 (Automorfismo en el toro.)**.** Consideremos:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

Observamos que $\det A = 1$. Esto implica que $A \in S(2, \mathbb{Z})$. Por lo tanto,

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}.$$

Con $\det A^{-1} = 1$. Esto implica que $A^{-1} \in S(2, \mathbb{Z})$.

Ahora definiremos lo que vamos a entender como la *medida de Lebesgue en el toro*. Sea m la medida de Lebesgue en $\mathbb{R}^n$ y sea $B \subseteq [0,1]^n$ un conjunto medible. Definimos:

$$[B] = \{[x] : x \in B\}.$$

Definimos la *medida de Lebesgue en el Toro* λ como,

$$\lambda([B]) = m(B).$$

A continuación probaremos que todo endomorfismo en el toro T_A preserva la medida de Lebesgue en el toro λ.

Proposición 8. *Cualquier endomorfismo en el toro preserva la medida de Lebesgue en el toro.*

Demostración. Sea $A \in G(n, \mathbb{Z})$, $r > 0$ y $x \in \mathbb{T}^n$. Se sabe que cada punto $x \in \mathbb{T}^n$, tiene un número $k = |\det A|$ de preimágenes bajo el endomorfismo en el toro T_A. Así pues,

$$(T_A)^{-1}(B_r(x)) = B_1 \cup \cdots \cup B_k,$$

donde $B_1, B_2, \cdots, B_k$ son k componentes conexas. Consideremos las inversas locales,

$$S_i : B_r(x) \to B_i \text{ de } T_A \qquad \forall i = 1, 2, \cdots, k.$$

Así pues, $T_A \circ S_i = Id$. Entonces $\mathrm{d}_y S_i = A^{-1}$ para toda $y \in B_r(x)$ y para cada $i = 1, 2, \cdots k$.

$$\begin{aligned}
\lambda((T_A)^{-1}(B_r(x))) &= \sum_{i=1}^{k} \lambda(B_i) \\
&= \sum_{i=1}^{k} \lambda(S_i(B_r(x))) \\
&= \sum_{i=1}^{k} \int_{B_r(x)} |\det \mathrm{d}y S_i| \mathrm{d}\lambda(x) \\
&= |\det A| \int_{B_r(x)} |\det A^{-1}| \mathrm{d}\lambda \\
&= \lambda(B_r(x)).
\end{aligned}$$

□

1.2.5. Cadenas de Markov

Las cadenas de Markov tienen gran importancia en diversas áreas de las matemáticas. En el estudio de los sistemas dinámicos son importantes ya que nos permiten *codificar* y extraer las propiedades de otras transformaciones que preservan la medida de una manera más sencilla. Así también, son importantes en el estudio de la Teoría de la Información y en Probabilidad como veremos mas adelante en el capítulo 5.

La información se suele representar como una secuencia discreta de símbolos, por ejemplo, un periódico, un libro o este mismo trabajo consiste de una secuencia muy larga de símbolos, o bien, una computadora organiza la información en bloques de 0's y 1's que pueden representar cosas diversas como música, etc. Así pues, vamos a considerar un conjunto de $k \in \mathbb{N}$ símbolos $\mathcal{A}_k$ llamado *alfabeto*. Cada símbolo en nuestro alfabeto $\mathcal{A}_k$ recibe el nombre de *letra*. Las secuencias finitas de letras reciben el nombre de *palabras* o *bloques*.

Las sucesiones símbolos en la vida real suelen ser finitas. Sin embargo, es útil trabajar con sucesiones infinitas. Podemos considerar una *fuente* o máquina que transmite símbolos a partir de un tiempo fijo sin interrupción. En este caso consideramos sucesiones infinitas en un sólo sentido.

$$(x_n)_{n\in\mathbb{N}} = (x_0 x_1 x_2 \cdots).$$

Una sucesión de letras infinita en ambos lados puede representar la historia completa de una *fuente* que ha transmitido letras por cada unidad de tiempo. Por conveniencia utilizaremos un punto para identificar el tiempo cero o el instante actual.

$$(x_n)_{n\in\mathbb{Z}} = (\cdots x_{-1}.x_0 x_1 x_2 \cdots).$$

Denotaremos al conjunto de todas las sucesiones infinitas de símbolos en un sentido que contienen símbolos del alfabeto $\mathcal{A}_k$ como mostramos a continuación y lo llamaremos el conjunto de las *sucesiones laterales*.

$$\Sigma_k^+ = \mathcal{A}_k^{\mathbb{N}}.$$

Al conjuntos de todas las sucesiones infinitas en ambos sentidos de símbolos en un alfabeto $\mathcal{A}_k$ lo llamaremos el conjuntos de las *sucesiones bilaterales* y lo denotaremos como sigue.

$$\Sigma_k = \mathcal{A}_k^{\mathbb{Z}}.$$

Decimos que una sucesión $x = (x_n)$ ya sea lateral o bilateral contiene un bloque o *palabra* $w = w_1 w_2 \cdots w_l$ si existe un j tal que $w_i = x_{j+i}$ para $i = 1, 2, \cdots, l$. Usualmente en los lenguajes no cualquier bloque de símbolos tiene sentido. Para restringir cuales bloques están permitidos y cuáles no, recurriremos las llamadas *matrices adyacentes*. Los subespacios resultantes reciben el nombre de "*shifts*"de Markov.

Una matriz adyacente es una matriz que pertenece al espacio $\mathcal{M}_{k\times k}(\{0,1\})$, es decir, son las matrices cuadradas $A = (a_{ij})$ con entradas de ceros y unos. Para explicar cómo es que la matriz adyacente describe las sucesiones que pertenecen al subespacio y cuáles no, vamos a considerar un alfabeto $\mathcal{A}_k$ y una numeración de sus elementos $\{1, 2, \cdots, k\}$ de manera que nos referiremos a los elementos del alfabeto por su número asignado. Cada entrada de la matriz nos indicará si puede ocurrir un determinado símbolo después de otro, dependiendo de si el valor correspondiente de la entrada en 1 o 0. Por ejemplo, si quiero saber si después del símbolo 2 puede ocurrir el símbolo 3, debo fijarme en la entrada de la matriz a_{23}. Si la entrada vale 1, significa que puede ocurrir. Si la entrada vale 0, significa que no puede ocurrir. Una manera más formal de expresar esta idea sería la siguiente:

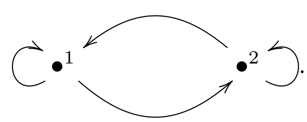
$$\Sigma_A^+ = \{(x_n)_{n\in\mathbb{N}} \in \Sigma_k^+ : a_{x_n x_{n+1}} = 1 \forall n \in \mathbb{N}\},$$

$$\Sigma_A = \{(x_n)_{n\in\mathbb{Z}} \in \Sigma_k : a_{x_n x_{n+1}} = 1 \forall n \in \mathbb{Z}\}.$$

A cada matriz adyacente A le podemos asociar una *gráfica dirigida* Γ_A que describa su comportamiento. Cada vértice de la gráfica representa los distintos símbolos del alfabeto. Las flechas que conecten los distintos vértices indicarán si después de un símbolo puede ocurrir otro.

Ejemplo 3. Consideremos una matriz es una matriz adyacente junto con su respectiva gráfica.

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$



Algunas de las transiciones posibles son,

$$1 \to 1 \to 1 \to 1 \to 1 \to 1.$$

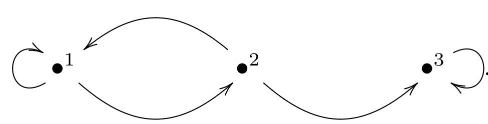
o bien,

$$1 \to 2 \to 1 \to 2 \to 2 \to 2.$$

Así pues, se pueden describir sucesiones como $(111111\cdots)$ o como $(\cdots 121222\cdots)$.

Ejemplo 4. Consideremos la siguiente matriz adyacente de 3×3 junto con su respectiva gráfica.

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$



Una transición posible sería,

$$1 \to 2 \to 1 \to 2 \to 3 \to 3 \to \cdots.$$

Sin embargo, una transición que no se admite es,

$$1 \to 2 \to 3 \to 2 \to 2.$$

Así pues, mientras la sucesión $(12123333\cdots)$ pertenece a Σ_A^+, la sucesión $(12322\cdots)$ no pertenece a Σ_A^+.

Dada una matriz adyacente $A = (a_{ij})$ con $i, j \in \{1, 2, \cdots k\}$, decimos que una palabra $w = w_1 w_2 \cdots w_l$ con $w_1, w_2, \cdots, w_l \in \mathcal{A}_k$ está *permitida* si $a_{x_n x_{n+1}} = 1$ para toda $n \in \mathbb{N}$, o bien, de manera equivalente, si existe una flecha en la gráfica Γ_A que va de x_n a x_{n+1}. Si una palabra no está permitida, decimos que está *prohibida*.

A continuación vamos a definir la *transformación corrimiento*. La transformación corrimiento se define sobre los *shifts* de Markov y consiste en recorrer la sucesión completa un paso a la izquierda. Esta transformación se puede interpretar como el paso del tiempo en una fuente que va emitiendo símbolos.

Definición 6. *Sea $A \in \mathcal{M}_{k\times k}(\{0,1\})$ una matriz adyacente y sea Σ_A^+ un shift de Markov lateral. Definimos la transformación **corrimiento lateral** $\sigma\colon \Sigma_A^+ \to \Sigma_A^+$ como la sucesión que se obtiene al recorrer un paso a la izquierda una sucesión dada.*

$$\sigma(x_0 x_1 x_2 \cdots) = (x_1 x_2 x_3 \cdots).$$

De manera análoga, se define la transformación corrimiento para sucesiones bilaterales.

Definición 7. *Sea $A \in \mathcal{M}_{k\times k}(\{0,1\})$ una matriz adyacente y sea Σ_A un shift de Markov bilateral. Definimos la transformación **corrimiento bilateral** $\sigma\colon \Sigma_A \to \Sigma_A$ como la sucesión que se obtiene al recorrer un paso a la izquierda una sucesión dada.*

$$\sigma(\cdots x_{-1}.x_0 x_1 x_2 \cdots) = (\ldots x_{-1} x_0 . x_1 x_2 \cdots).$$

Observamos que de la definición se sigue que la transformación corrimiento es invertible cuando se aplica a sucesiones bilaterales, pero no es invertible cuando se aplica sobre sucesiones laterales.

A continuación vamos a definir dos medidas, a saber, la medida de Bernoulli y la medida de Markov. Posteriormente probaremos que la transformación corrimiento preserva ambas medidas. Para definir la medida de Bernoulli y la medida de Markov, definiremos antes una σ-álgebra adecuada generada a partir de los *cilindros*. Definiremos primero los cilindros en el espacio Σ_k^+.

Definición 8. *Sea $\mathcal{A}_k$ un alfabeto, sea Σ_k^+ el espacio de sucesiones laterales, $m \in \mathbb{N}$ y $i_1, \cdots, i_m \in \mathcal{A}_k$. Definimos un **cilindro** $C_{i_1,\cdots,i_m} \subseteq \Sigma_k^+$ como:*

$$C_{i_1,\cdots,i_m} = \{(j_1 j_2 \cdots) \in \Sigma_k^+ : j_l = i_l \forall l = 1, \cdots, m\}.$$

Así también, de manera análoga, definimos los cilindros para el espacio de sucesiones bilaterales Σ_k.

Definición 9. *Sea $\mathcal{A}_k$ un alfabeto, Σ_k el espacio de sucesiones bilaterales, $r \in \mathbb{Z}$, $m \in \mathbb{N}$ y $i_r, i_{r+1} \cdots, i_{r+m} \in \mathcal{A}_k$. Definimos un **cilindro** $C_{i_r, i_{r+1} \cdots, i_{r+m}} \subseteq \Sigma_k$ como*

$$C_{i_r, i_{r+1} \cdots, i_{r+m}} = \{(\cdots j_s j_{s+1} j_{s+2} \cdots) \in \Sigma_k : j_l = i_l \forall l = r, r+1, \cdots, r+m\}.$$

Un tipo de cilindro particularmente importante son los llamados *cilindros centrados en el origen* que se definen como sigue.

Definición 10. *Sea Σ_k el espacio de sucesiones bilaterales. Sea $m \in \mathbb{N}$ y $i_{-m}, \cdots, i_m \in \mathcal{A}_k$. Definimos un* ***cilindro centrado en el origen*** $C_{i_{-m},\cdots,i_m} \subseteq \Sigma_k$ *como*

$$C_{i_{-m},\cdots,i_m} = \{(\cdots j_{-1} j_0 j_1 \cdots) \in \Sigma_k : j_l = i_l \forall l = -m, \cdots, m\}.$$

Tanto en el espacio Σ_k como en el espacio Σ_k^+, la σ-álgebra con la que trabajaremos es la σ-álgebra generada por los respectivos cilindros. Para definir la medida de Bernoulli definiremos primero un *vector de probabilidad*.

Definición 11. *Sea $k \in \mathbb{N}$ y $p = (p_1, p_2, \cdots, p_k) \in (0,1)^k$. Decimos que p es un* ***vector de probabilidad*** *si*

$$\sum_{i=1}^{k} p_i = 1.$$

Con base en un vector de probabilidad, definimos la medida de Bernoulli lateral y la medida de Bernoulli bilateral de manera explícita sobre los cilindros que son generadores de la σ-álgebra S.

Definición 12. *Sea (Σ_k^+, S) un espacio medible y $p = (p_1, p_2, \cdots, p_k)$ un vector de probabilidad. Definimos la* ***medida de Bernoulli lateral*** *como*

$$\mu_p(C_{i_1 i_2 \cdots i_m}) = p_{i_1} \cdots p_{i_m}.$$

Para cada $m \in \mathbb{N}$ y $i_1, \cdots, i_m \in \{1, \cdots, k\}$.

Así también, de manera análoga, definimos la medida de Bernoulli en el espacio Σ_k.

Definición 13. *Sea (Σ_k, S) un espacio medible y $p = (p_1, p_2, \cdots, p_k)$ un vector de probabilidad. Definimos la* ***medida de Bernoulli bilateral*** *como*

$$\mu_p(C_{i_r i_{r+1} \cdots i_{r+m}}) = p_{i_r} p_{i_{r+1}} \cdots p_{i_{r+m}}.$$

Para cada $r \in \mathbb{Z}$, $m \in \mathbb{N}$ y $i_r, i_{r+1} \cdots, i_{r+m} \in \{1, \cdots k\}$.

Ejemplo 5. Consideremos el vector

$$p = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right).$$

Observamos que el vector p es un vector de probabilidad ya que la suma de sus entradas es 1. Calculamos la medida de Bernoulli para el cilindro C_{1132},

$$\mu_p(C_{1132}) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{72}.$$

Para definir ahora la medida de Markov, definiremos primero una *matriz estocástica*

Definición 14. *Sea $k \in \mathbb{N}$ y sea $P = (p_{ij})$ una matriz cuadrada $k \times k$ tal que $p_{ij} \geq 0$ para cada $i, j = 1, 2, \cdots, k$. Decimos que P es una* ***matriz estocástica*** *si la suma de sus filas resulta 1, es decir,*

$$\sum_{j=1}^{k} p_{ij} = 1 \qquad \forall i = 1, 2, \cdots k.$$

Para definir la medida de Markov, necesitamos asociar a las matrices estocásticas un vector de probabilidad. Decimos que una matriz estocástica $P = (p_{ij})$ y un vector de probabilidad $p = (p_1, p_2, \cdots, p_k)$ forman un ***par estocástico*** (P, p) si se cumple $p = Pp$, es decir,

$$\sum_{i=1}^{k} p_i p_{ij} = p_j \qquad \forall j = 1, 2, \cdots, k.$$

No siempre podemos dada una matriz estocástica P asociarle un vector de probabilidad p tal que (P, p) forme un par estocástico. El siguiente teorema muestra las condiciones que debe cumplir la matriz estocástica para que pueda existir un par estocástico. Mas aún, se prueba que el vector de probabilidad asociado a la matriz estocástica para formar un par estocástico es único.

Teorema 4 (Perron-Frobenius)**.** *Sea $P = (p_{ij})$ con $i, j = 1, 2, \cdots, k$ una matriz estocástica. Si P es una matriz irreducible, es decir, que para cada entrada existe un $m \in \mathbb{N}$ tal que la entrada ij de la matriz P^m es positiva, entonces existe un único vector de probabilidad p tal que (P, p) es un par estocástico.*

Demostración. Consideremos la norma $\|v\| = \sum_{i=1}^{k} |v_i|$ y definimos el conjunto,

$$S = \{v \in (\mathbb{R}_0^+)^k : \|v\| = 1\}.$$

Así también, definimos $F \colon S \to S$ como

$$F(v) = \frac{Pv}{\|Pv\|}.$$

Observamos que la función F es continua. Sabemos que S es homeomorfa a la bola cerrada unitaria en $\mathbb{R}^{k-1}$. Por el Teorema del punto fijo de Brouwer sabemos que existe $p \in S$ tal que,

$$p = F(p) = \frac{Pp}{\|Pp\|}.$$

Ahora bien, observamos que por definición de la norma que estamos utilizando

$$\begin{aligned}\|Pp\| &= \sum_{j=1}^{k} |(Pp)j| \\ &= \sum_{j=1}^{k}\sum_{i=1}^{k} |p_i|\, p_{ij} \\ &= \sum_{i=1}^{k} |p_i| \sum_{j=1}^{k} p_{ij} \\ &= \sum_{i=1}^{k} |p_i| \\ &= \|p\| = 1.\end{aligned}$$

Veamos que las entradas del vector p son no negativas. Como $p \in S$ debe tener al menos una entrada positiva, digamos p_j para $j \in \{1, 2, \cdots, l\}$. Sea $i \in \{1, 2, \cdots, k\}$. Por hipótesis, P es una matriz irreducible, entonces existe $m \in \mathbb{N}$ tal que la entrada ij de la matriz P^m es positiva. Como p es no negativo y $p = Pp$ se tiene que

$$p_i = \sum_{l=1}^{k} (P^m)_{il} p_l \geq (P^m)_{ij} p_j > 0.$$

Así pues, las entradas de p son necesariamente positivas. Veamos que el vector p es único. Supongamos que existe $q = (q_1, q_2, \cdots, q_k) \in S$ diferente de p con entradas positivas y que además cumple que $Pq = q$. Consideremos

$$t = \text{mín}\left\{\frac{p_1}{q_1}, \frac{p_2}{q_2}, \cdots, \frac{p_k}{q_k}\right\}$$

y nos fijamos en el vector $p - tq$. Como $t = \frac{p_j}{q_j}$ para alguna $j \in \{1, 2, \cdots, k\}$, entonces la entrada $p_j - tq_j = 0$ mientras que las demás entradas son mayores o iguales a cero. En efecto, supongamos que $l \neq j$. Como

$$\frac{p_l}{q_l} \geq \frac{p_j}{q_j}$$

esto implica que

$$p_l - \frac{p_j}{q_j} q_l = p_l - tq_l \geq 0.$$

Las entradas del vector $p - tq$ son mayores o iguales a cero salvo la entrada j que es igual a 0. Observamos también que $p - tq$ es un vector característico, pues

$$P(p - tq) = Pp - tPq = p - tq.$$

Sin embargo, dado que por hipótesis tenemos que P es una matriz irreducible, por un argumento análogo al que se utilizó para probar que las entradas de

p son positivas, se puede probar que las entradas de $p - tq$ son necesariamente positivas. Lo cual es una contradicción ya que la entrada j vale 0. Por lo tanto, el vector p es único. $\square$

A partir de un par estocástico (P, p), vamos a definir la medida de Markov lateral y la medida de Markov bilateral. Recordamos que la σ-álgebra S sobre la que definiremos las medidas es la σ-álgebra generada por los cilindros respectivos.

Definición 15. *Sea (σ_k, S) un espacio medible y sea (P, p) un par estocástico, definimos la* ***medida de Markov lateral*** *asociada al par estocástico (P, p) como*

$$\mu_{(P,p)}(C_{i_1 \cdots i_m}) = p_i p_{i_1 i_2} \cdots p_{i_{m-1} i_m}.$$

Para cada $m \in \mathbb{N}$ y $i_1, i_2, \cdots, i_m \in \{1, 2, \cdots k\}$.

Así también, de manera análoga, definimos la medida de Markov en el espacio Σ_k.

Definición 16. *Sea (Σ_k, S) un espacio medible y sea (P, p) un par estocástico. Definimos la* ***medida de Markov bilateral*** *asociada al par estocástico (P, p) como,*

$$\mu_{(P,p)}(C_{i_r i_{r+1} \cdots i_{r+m}}) = p_{i_r} p_{i_r i_{r+1}} p_{i_{r+1} i_{r+2}} \cdots p_{i_{r+m-1} i_{r+m}}.$$

Para cada $r \in \mathbb{Z}$, $m \in \mathbb{N}$ y $i_r i_{r+1} \cdots i_{r+m} \in \{1, \cdots k\}$.

Las medidas de Markov están bien definidas porque los cilindros son generadores de la σ-álgebra S.

Ejemplo 6. Consideremos la matriz P

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}.$$

Observamos que la matriz P es una matriz estocástica, ya que sus filas suman 1. Así también, observamos que el vector de probabilidad $p = (1/2, 1/2)$ es un vector característico y positivo de la matriz P y entonces $p = pP$. Así pues, se tiene que (P, p) es un par estocástico. Calculamos la medida de Markov para el cilindro C_{121},

$$\mu_{(P,p)}(C_{121}) = p_1 p_{12} p_{21} = \frac{1}{2} \cdot \frac{1}{2} \cdot 1 = \frac{1}{4}.$$

Antes de probar que la transformación corrimiento preserva tanto la medida de Bernoulli como la medida de Markov, es importante mencionar que la medida de Bernoulli se puede ver como un caso particular de la medida de Markov. Basta tomar $p_{ij} = p_j$ para cada i y j como la matriz estocástica. Así pues, sólo necesitamos probar que σ preserva la medida de Markov.

Proposición 9. *Sea (Σ_k, S, μ) un espacio de medida donde S es la σ-álgebra generada por los conjuntos cilindros y μ la medida de Markov bilateral. Entonces, la transformación corrimiento σ preserva la medida de Markov bilateral.*

Demostración. Sea $\mu_{(P,p)}$ la medida de Markov asociada al par estocástico (P,p). Sea $r \in \mathbb{Z}$ y $m \in \mathbb{N}$. Entonces

$$\sigma^{-1}(C_{i_r i_{r+1} \cdots i_{r+m}}) = C_{i_r i_{r+1} \cdots i_{r+m}}.$$

Así pues, es inmediato que

$$\mu_{(P,p)}(\sigma^{-1}(C_{i_r i_{r+1} \cdots i_{r+m}})) = \mu_{(P,p)}(C_{i_r i_{r+1} \cdots i_{r+m}}).$$

Por lo tanto, σ preserva la medida $\mu_{(P,p)}$.

□

Observamos que en particular la transformación σ preserva la medida para los cilindros centrados en el origen. A continuación vamos a probar que la transformación σ preserva la medida de Markov lateral.

Proposición 10. *Sea (Σ_k^+, S, μ) un espacio de medida donde S es la σ-álgebra generada por los cilindros y μ la medida de Markov lateral. Entonces, la transformación corrimiento σ preserva la medida de Markov lateral.*

Demostración. Sea $\mu_{(P,p)}$ una medida de Markov asociada a un par estocástico (P,p). Se tiene lo siguiente

$$\sigma^{-1}(C_{i_1 \cdots i_m}) = \bigcup_{j=1}^{k} C_{j i_1 \cdots i_m}.$$

Observamos que $\bigcup_{j=1}^{k} C_{j i_1 \cdots i_m}$ es una unión ajena. Así pues, se tiene que

$$\begin{aligned}
\mu(\sigma^{-1}(C_{i_1 \cdots i_m})) &= \mu\left(\bigcup_{j=1}^{k} C_{j i_1 \cdots i_m}\right) \\
&= \sum_{j=1}^{k} \mu(C_{j i_1 \cdots i_m}) \\
&= \sum_{j=1}^{k} p_j p_{j i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m} p_{i_m j} \\
&= p_{i_1 i_2} \cdots p_{i_{m-1} i_m} \sum_{j=1}^{k} p_j p_{j i_1} \\
&= p_{i_1} p_{i_1 i_2} \cdots p_{i_{m-1} i_m} \\
&= \mu(C_{i_1 \cdots i_m}).
\end{aligned}$$

Por lo tanto, σ preserva la medida de Markov. □

La medida de Markov y la medida de Bernoulli son de suma importancia en Probabilidad. Podemos considerar un conjunto finito de eventos mutuamente exclusivos con sus respectivas probabilidades, los eventos pueden ser dependientes o independientes. Por una parte, si consideramos que los eventos son independientes, entonces se puede calcular la probabilidad de que ocurran eventos simultáneos al calcular su medida de Bernoulli. Por otra parte, si consideramos que los eventos son mutuamente dependientes, entonces la probabilidad de que ocurra un evento después de que ocurrió otro, depende del valor correspondiente en su matriz de transición y de la probabilidad del estado. Se puede calcular la probabilidad de un determinado estado al calcular su medida de Markov.

1.3. Existencia de las Medidas Invariantes

En la sección anterior 1.2 se presentaron algunos de los ejemplos más representativos de transformaciones que preservan una medida para espacios de medida finita. Sin embargo, no es claro que dada una transformación medible T definida en un espacio de medida (X, S, μ'), siempre se pueda encontrar una medida μ no trivial, definida sobre la misma σ-álgebra S, de manera que T preserve la medida μ.

En esta sección mostraremos que el *problema de la existencia de las medidas invariantes* tiene solución al menos para *espacios métricos compactos* y cuando la transformación T es *continua*. En la primera parte, mostraremos algunas de las propiedades principales del *espacio de medidas finitas* $\mathcal{M}(X)$ y posteriormente presentaremos el *Teorema de Kyrilov-Bogolubov* que condensa la solución al problema de la existencia de medidas invariantes.

1.3.1. Espacio de Medidas

A continuación probaremos algunas de las propiedades principales que tienen los espacios de medida finitos. Dado un espacio de medida (X, S, μ'), denotamos por $\mathcal{M}(X)$ al espacio de medidas definidas sobre la misma σ-álgebra tales que $\mu(X) = 1$. Mostraremos que si el espacio X es un espacio métrico compacto, entonces el espacio de medidas $\mathcal{M}(X)$ es un espacio métrico, compacto y convexo.

El espacio de medidas es un espacio métrico.

Para mostrar que el espacio de medidas $\mathcal{M}(X)$ es un espacio métrico, se definirá de manera explícita una métrica d. Para definir a la métrica d, probaremos antes algunos resultados auxiliares. Sea (X, d) un espacio métrico compacto. Denotamos como $C(X)$ al conjunto de funciones continuas con valores reales.

$$C(X) = \{\phi : X \to \mathbb{R} : \phi \text{ es continua}\}.$$

Observamos que el conjunto $C(X)$ tiene estructura de espacio vectorial con las operaciones:

$$(f + g)(x) = f(x) + g(x)$$
$$(\lambda f)(x) = \lambda(f(x)).$$

Más aun, definimos la norma $\|\cdot\|_\infty$ sobre $C(X)$ como sigue:

$$\|\phi\|_\infty = \text{máx}\{|\phi(x)| : x \in X\}.$$

A continuación probaremos que el espacio $(C(X), \|\cdot\|_\infty)$ es separable, es decir, que tiene un conjunto denso y numerable. La prueba del teorema ocupa el Teorema de Stone-Weierstrass, recordaremos lo que dice el teorema y la definición de álgebra de funciones.

Definición 17. *Sea X un espacio métrico compacto. Decimos que $D \subseteq C(X)$ es un **álgebra de funciones** si para $f, g \in D$ y $\lambda \in \mathbb{R}$ se cumple lo siguiente:*

- $f + g \in D$.
- $fg \in D$.
- $\lambda f \in D$.

El Teorema de Stone-Weierstrass se enuncia en términos de familias de funciones que son *separadoras de puntos*, es decir, para cualesquiera dos puntos, podemos encontrar al menos dos funciones en la familia tales que los valores de los puntos evaluados son distintos. El Teorema de Stone-Weierstrass afirma que si una familia de funciones es separadora, entonces la cerradura del álgebra de funciones más pequeña que la contiene coincide con el conjunto de funciones continuas.

Teorema 5 (Teorema de Stone-Weierstrass)**.** *Sea X un espacio compacto y T_2[2]. Si $D \subseteq C(X)$ es una familia de funciones que separa los puntos de X y contiene a la función constante $\mathbf{1}$, entonces la cerradura del álgebra generada por la familia D es idéntica con $C(X)$.*

Para probar que el espacio de funciones continuas es separable, definiremos una familia de funciones continuas numerable con estructura de álgebra que satisfaga las hipótesis del Teorema de Stone-Weierstrass, es decir, que sea una familia separadora. De esta manera, podremos afirmar que la familia definida es ella misma un conjunto denso en el espacio de funciones continuas.

Teorema 6. *Sea (X, ρ) un espacio métrico compacto. Entonces, el espacio $(C(X), \|\cdot\|_\infty)$ es separable.*

Demostración. Por hipótesis tenemos que el espacio X es un espacio métrico compacto, entonces se tiene que X es un espacio *separable*, es decir, existe un conjunto $D \subseteq X$ tal que es *denso* y *numerable*. Por el *principio del buen orden*, sabemos que podemos numerar a los elementos del conjunto D. Supongamos que el conjunto $(x_n)_{n\in\mathbb{N}}$ es una numeración del conjunto D. A partir del conjunto

[2]Es decir, satisface el siguiente axioma de separación: $\forall x, y \in X, \exists U, V$ abiertos tales que $x \in U, y \in V$ y $U \cap V = \emptyset$

$D = (x_n)_{n\in\mathbb{N}}$ definimos la familia de funciones $\{\chi_{nm}\}_{n,m\in\mathbb{N}}$ como sigue, para cada $n, m \in \mathbb{N}$,

$$\chi_{nm}(x) = \begin{cases} \dfrac{1}{m} - \rho(x, x_n) & \text{si } \rho(x, x_n) \leq \dfrac{1}{m} \\ 0 & \text{si } \rho(x, x_n) > \dfrac{1}{m}. \end{cases}$$

Vamos a probar que la familia $\{\chi_{nm}\}_{n,m\in\mathbb{N}}$ satisface las condiciones del Teorema de Stone-Weierstrass, es decir, probaremos que las funciones de la familia son continuas y que la familia separa puntos.

Afirmación: Las funciones de la familia $\{\chi_{nm}\}_{n,m\in\mathbb{N}}$ son continuas.

Sean $n, m \in \mathbb{N}$ y sea $\epsilon > 0$. Tomamos $\delta = \epsilon$ y supongamos que $\rho(x, y) < \delta$. Veamos que $|\chi_{nm}(x) - \chi_{nm}(y)| < \epsilon$, para ello, analizamos los siguientes casos:

Caso 1: $\rho(x, x_n) \leq \frac{1}{m}$ y $\rho(y, x_n) \leq \frac{1}{m}$. Entonces se sigue que,

$$\begin{aligned} |\chi_{nm}(x) - \chi_{nm}(y)| &= \left| \frac{1}{m} - \rho(x, x_n) - \frac{1}{m} - \rho(y, x_n) \right| \\ &= |\rho(x, x_n) - \rho(y, x_n)| \\ &\leq \rho(x, y) < \delta = \epsilon. \end{aligned}$$

Caso 2: $\rho(x, x_n) > \frac{1}{m}$ y $\rho(y, x_n) > \frac{1}{m}$. Luego,

$$|\chi_{nm}(x) - \chi_{nm}(y)| = |0 - 0| < \epsilon.$$

Caso 3: $\rho(x, x_n) \leq \frac{1}{m}$ y $\rho(y, x_n) > \frac{1}{m}$. Esto implica,

$$|\chi_{nm}(x) - \chi_{nm}(y)| = \left| \frac{1}{m} - \rho(x, x_n) - 0 \right|.$$

Como $\rho(x, x_n) \leq \frac{1}{m} < \rho(y, x_n)$, entonces,

$$0 \leq \frac{1}{m} - \rho(x, x_n) < \rho(y, x_n) - \rho(x, x_n) \leq \rho(x, y) < \delta = \epsilon.$$

Por lo tanto, en cualquier caso, χ_{nm} es continua para cada $n, m \in \mathbb{N}$. Así también, veamos que cumple la segunda condición de Teorema de Stone-Weierstrass.

Afirmación: La familia de funciones $\{\chi_{nm}\}_{n,m\in\mathbb{N}}$ separa puntos, es decir, para cada $x, y \in X$, existen $n, m \in \mathbb{N}$ tales que $\chi_{nm}(x) \neq \chi_{nm}(y)$. Sean $x, y \in X$. Consideramos $\rho(x, y) = \delta$ y nos fijamos en una $N \in \mathbb{N}$ lo suficientemente grande de tal manera que,

$$\left(\frac{N}{(N-1)\delta}, \frac{N}{\delta} \right) \cap \mathbb{N} \neq \emptyset.$$

Sea $m \in \left(\frac{N}{(N-1)\delta}, \frac{N}{\delta} \right) \cap \mathbb{N}$. Entonces se tiene,

$$\frac{N}{(N-1)\delta} < m < \frac{N}{\delta}.$$

Esto implica que,

$$\frac{\delta}{N} < \frac{1}{m} < \frac{(N-1)\delta}{N}.$$

Consideramos $B_{\frac{\delta}{N}}(x)$. Por densidad de D, se sigue que existe $n \in \mathbb{N}$ tal que,

$$x_n \in B_{\frac{\delta}{N}}(x) \cap D.$$

Luego,

$$\rho(x_n, x) \leq \frac{\delta}{N} < \frac{1}{m}.$$

Por definición se tiene que $\chi_{nm}(x) = \frac{1}{m} - \rho(x_n, x) > 0$. Veamos que $\frac{1}{m} < \rho(y, x_n)$. Como,

$$\rho(x_n, x) \leq \frac{\delta}{N} = \frac{\rho(x,y)}{N},$$

se sigue

$$-\frac{\rho(x,y)}{N} \leq -\rho(x_n, x).$$

Sumando el término $\rho(x,y)$ a ambos lados de la desigualdad y por la desigualdad del triángulo

$$\rho(x,y) - \frac{\rho(x,y)}{N} = \delta - \frac{\delta}{N} = \frac{(N-1)\delta}{N} \leq \rho(x,y) - \rho(x_n, x) \leq \rho(y, x_n),$$

pero,

$$\frac{1}{m} < \frac{(N-1)\delta}{N} \leq \rho(y, x_n).$$

Entonces, se tiene que $\chi_{nm}(y) = 0$. Por lo tanto, se tiene que $\chi_{nm}(x) \neq \chi_{nm}(y)$ y entonces concluimos que la familia $\{\chi_{nm}\}_{n,m\in\mathbb{N}}$ separa puntos. Si consideramos la familia de funciones $K = \{\chi_{nm}\}_{n,m\in\mathbb{N}} \cup \{1\}$, donde 1 es la función constante 1. Por el Teorema de Stone-Weierstrass se tiene que,

$$\overline{\mathcal{A}(K)} = C(X),$$

donde $\mathcal{A}(K)$ es el álgebra generada por la funciones de la familia K. Observamos que el conjunto $\mathcal{A}(K)$ es un conjunto denso sobre $C(X)$. Sin embargo, no se garantiza que sea un conjunto numerable.

Definimos la familia de funciones $\mathcal{B}$ tal que para cada $N \in \mathbb{N}$ tienen la forma,

$$c_0 + \sum_{i=1}^{N} c_i \chi_{n_i m_i},$$

donde $c_0, c_1, \cdots, c_N \in \mathbb{Q}$, $n_i, m_i \in \mathbb{N}\ \forall i \in \{1, 2, \cdots, N\}$ y sus productos finitos. Como, por definición de álgebra de funciones, los elementos de $\mathcal{A}(K')$ son de la forma

$$c_0 + \sum_{i=1}^{N} c_i \chi_{n_i m_i},$$

donde $c_0, c_1, \cdots, c_N \in \mathbb{R}$, $n_i, m_i \in \mathbb{N}$ $\forall i \in \{1, 2, \cdots, N\}$ y sus productos finitos, entonces se tiene que,

$$\mathcal{B} \subseteq \mathcal{A}(K').$$

Más aún, como sabemos que cada número real se puede aproximar por una sucesión de números racionales, entonces tenemos que,

$$\overline{\mathcal{B}} = \mathcal{A}(K')$$

y como la cardinalidad de $\mathbb{Q}$ es $\aleph_0$, se concluye que la cardinalidad de $\mathcal{B}$ es $\aleph_0$. Por lo tanto, $\mathcal{B}$ es un conjunto denso y numerable en $C(X)$ y así concluimos que $C(X)$ es separable. □

El Teorema 6 asegura que dado un espacio métrico (X, ρ), el conjunto de funciones continuas $C(X)$ es separable, es decir, existe un conjunto $D \subseteq C(X)$ denso y numerable. Para definir una métrica sobre el espacio de medidas finitas $\mathcal{M}(X)$, nos restringiremos a las funciones en $C(X)$ tales que $\|\phi\|_\infty \leq 1$. Definimos el siguiente conjunto,

$$H = \{\phi \in C(X) : ||\phi||_\infty \leq 1\} \cap D. \tag{1.5}$$

Observamos que el conjunto H es también denso y numerable. Por el principio del buen orden, podemos considerar una numeración de los elementos de H, digamos $(\phi_n)_{n\in\mathbb{N}}$. Definimos $d : \mathcal{M}(X) \times \mathcal{M}(X) \to \mathbb{R}^+$ como sigue:

$$d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left| \int \phi_n \mathrm{d}\mu - \int \phi_n \mathrm{d}\nu \right|. \tag{1.6}$$

A continuación probaremos que la función d que se acaba de definir es una métrica para $\mathcal{M}(X)$. Primero probaremos que está bien definida y posteriormente probaremos que satisface las condiciones para ser considerada una métrica. El siguiente lema tiene como consecuencia directamente que d está bien definida.

Lema 1. *Sea (X, ρ) un espacio métrico, H definido como en (1.5) y $\mathcal{M}(X)$ el espacio de medidas finitas en X definidas sobre la σ-álgebra S. Entonces la función definida en (1.6) satisface la siguiente desigualdad:*

$$d(\mu, \nu) \leq \sum_{n=1}^{\infty} \frac{2||\phi_n||_\infty}{2^n} \leq 2 \qquad \forall \mu, \nu \in \mathcal{M}(X), \tag{1.7}$$

donde $\phi_n \in H$ para cada $n \in \mathbb{N}$.

Demostración. Por la desigualdad del triángulo y por propiedades de la integral se obtiene lo siguiente:

$$\begin{aligned}\left|\int \phi_n \mathrm{d}\mu - \int \phi_n \mathrm{d}\nu\right| &\leq \left|\int \phi_n \mathrm{d}\mu\right| + \left|\int \phi_n \mathrm{d}\nu\right| \qquad \forall n \in \mathbb{N}\\ &\leq \int |\phi_n| \mathrm{d}\mu + \int |\phi_n| \mathrm{d}\nu \qquad \forall n \in \mathbb{N}\\ &\leq \int \|\phi_n\|_\infty \mathrm{d}\mu + \int \|\phi_n\|_\infty \mathrm{d}\nu \qquad \forall n \in \mathbb{N}\\ &\leq \|\phi_n\|_\infty (\mu(X) + \nu(X)) \qquad \forall n \in \mathbb{N}\\ &= 2\|\phi_n\|_\infty \qquad \forall n \in \mathbb{N}.\end{aligned}$$

Como $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$ y $\|\phi_n\| \leq 1$ $\forall n \in \mathbb{N}$, entonces tenemos,

$$\begin{aligned}d(\mu, \nu) &\leq \sum_{n=1}^{\infty} \frac{2\|\phi_n\|_\infty}{2^n}\\ &\leq 2\sum_{n=1}^{\infty} \frac{1}{2^n} = 2.\end{aligned}$$

□

Por el Lema 1 se tiene que la suma infinita definida por la función $d : \mathcal{M}(X) \times \mathcal{M}(X) \to \mathbb{R}^+$ es una suma convergente para cada $\mu, \nu \in \mathcal{M}(X)$. Por lo tanto, la función d está bien definida. Veamos ahora que la función d satisface las propiedades de una métrica.

Proposición 11. *Sea (X, ρ) un espacio métrico compacto y H como en (1.5). La función $d : \mathcal{M}(X) \times \mathcal{M}(X) \to \mathbb{R}^+$ definida como:*

$$d(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} \left|\int \phi_n \mathrm{d}\mu - \int \phi_n \mathrm{d}\nu\right|$$

es una métrica para $\mathcal{M}(X)$, donde $\phi_n \in H$ para cada $n \in \mathbb{N}$.

Demostración. Para probar que d es una métrica para $\mathcal{M}(X)$ se debe verificar que se cumplen las siguientes propiedades:

- $d(\mu, \nu) \geq 0 \qquad \forall \mu, \nu \in \mathcal{M}(X)$.
- $d(\mu, \nu) = d(\nu, \mu) \qquad \forall \mu, \nu \in \mathcal{M}(X)$.
- $d(\mu, \nu) \leq d(\mu, \rho) + d(\rho, \nu) \qquad \forall \mu, \nu, \rho \in \mathcal{M}(X)$.
- $d(\mu, \nu) = 0 \Leftrightarrow \nu = \mu$.

Para la probar la primera propiedad, observamos que,

$$0 \leq \left|\int \phi_n \mathrm{d}\mu - \int \phi_n \mathrm{d}\nu\right| = c_n \qquad \forall n \in \mathbb{N}.$$

Esto implica que

$$d(\mu,\nu)=\sum_{n=1}^{\infty}\frac{1}{2^n}c_n\geq 0.$$

Veamos la segunda propiedad.

$$\begin{aligned}d(\mu,\nu)&=\sum_{n=1}^{\infty}\frac{1}{2^n}\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\nu\right|\\&=\sum_{n=1}^{\infty}\frac{1}{2^n}\left|\int\phi_n\mathrm{d}\nu-\int\phi_n\mathrm{d}\mu\right|\\&=d(\nu,\mu).\end{aligned}$$

La tercera propiedad.

$$\begin{aligned}\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\nu\right|&=\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\rho+\int\phi_n\mathrm{d}\rho-\int\phi_n\mathrm{d}\nu\right|\\&\leq\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\rho\right|+\left|\int\phi_n\mathrm{d}\rho-\int\phi_n\mathrm{d}\nu\right|.\end{aligned}$$

Entonces,

$$\begin{aligned}d(\mu,\nu)&\leq\sum_{n=1}^{\infty}\frac{1}{2^n}\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\rho\right|+\sum_{n=1}^{\infty}\frac{1}{2^n}\left|\int\phi_n\mathrm{d}\rho-\int\phi_n\mathrm{d}\nu\right|\\&=d(\mu,\rho)+d(\rho,\nu).\end{aligned}$$

Finalmente la cuarta propiedad. Supongamos primero que $\nu=\mu$. Esto implica que,

$$d(\mu,\nu)=d(\mu,\mu)=\sum_{n=1}^{\infty}\frac{1}{2^n}\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\mu\right|=0.$$

Ahora supongamos que $d(\mu,\nu)=0$ y veamos que $\mu=\nu$. Para probar que $\mu=\nu$ se probará que $\int\phi\mathrm{d}\mu=\int\phi\mathrm{d}\nu$ para toda función $\phi\in C(X)$. Así pues,

$$\sum_{n=1}^{\infty}\frac{1}{2^n}\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\nu\right|=0.$$

Esto implica que,

$$\left|\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\nu\right|=0\qquad\forall n\in\mathbb{N}.$$

Sin embargo, esto ocurre si y sólo si

$$\int\phi_n\mathrm{d}\mu-\int\phi_n\mathrm{d}\nu=0\qquad\forall n\in\mathbb{N}.$$

Entonces
$$\int \phi_n \mathrm{d}\mu = \int \phi_n \mathrm{d}\nu \qquad \forall n \in \mathbb{N}. \tag{1.8}$$
Ahora bien, si $\phi \in C(X)$, entonces
$$\frac{\phi}{\|\phi\|_\infty} \in \{\phi \in C(X) : \|\phi\|_\infty \leq 1\}.$$
Como $H = (\phi_n)_{n\in\mathbb{N}}$ es denso en $\{\phi \in C(X) : \|\phi\|_\infty \leq 1\}$, entonces, para $\epsilon \geq 0$, existe $m \in \mathbb{N}$ tal que,
$$\left\| \frac{\phi}{\|\phi\|_\infty} - \phi_m \right\|_\infty < \epsilon.$$
Como sabemos que,
$$\left| \frac{\phi(x)}{\|\phi\|_\infty} - \phi_m(x) \right| \leq \left\| \frac{\phi}{\|\phi\|_\infty} - \phi_m \right\|_\infty < \epsilon \qquad \forall x \in X,$$
se sigue que,
$$\left| \frac{\phi(x)}{\|\phi\|_\infty} - \phi_m(x) \right| < \epsilon \qquad \forall x \in X.$$
Integrando,
$$\left| \int \left(\frac{\phi(x)}{\|\phi\|_\infty} - \phi_m(x) \right) \mathrm{d}\mu \right| \leq \int \left| \frac{\phi(x)}{\|\phi\|_\infty} - \phi_m(x) \right| \mathrm{d}\mu \leq \mu(X)\epsilon = \epsilon.$$
Entonces se tiene
$$-\epsilon < \frac{1}{\|\phi\|_\infty} \int \phi \mathrm{d}\mu - \int \phi_m \mathrm{d}\mu < \epsilon.$$
Análogamente,
$$-\epsilon < -\frac{1}{\|\phi\|_\infty} \int \phi \mathrm{d}\nu + \int \phi_m \mathrm{d}\nu < \epsilon.$$
Sumando ambas desigualdades se obtiene
$$-2\epsilon < -\frac{1}{\|\phi\|_\infty} \left(\int \phi \mathrm{d}\mu - \int \phi \mathrm{d}\nu \right) < \int \phi_m \mathrm{d}\mu - \int \phi_m \mathrm{d}\nu + 2\epsilon.$$
Pero
$$\int \phi_m \mathrm{d}\mu - \int \phi_m \mathrm{d}\nu + 2\epsilon \leq \left| \int \phi_m \mathrm{d}\mu - \int \phi_m \mathrm{d}\nu \right| + 2\epsilon = 2\epsilon.$$
Pues habíamos visto en (1.8) que $\int \phi_m \mathrm{d}\mu = \int \phi_m \mathrm{d}\nu$. Por lo tanto,
$$-\frac{1}{\|\phi\|_\infty} \left| \int \phi \mathrm{d}\mu - \int \phi \mathrm{d}\nu \right| \leq 2\epsilon.$$
Como ϵ es arbitrario, se concluye que,
$$\int \phi \mathrm{d}\mu = \int \phi \mathrm{d}\nu \qquad \forall \phi \in C(X).$$

Por lo tanto $\mu = \nu$. Así pues, la función $d : \mathcal{M}(X) \times \mathcal{M}(X) \to \mathbb{R}^+$ es una métrica. $\square$

El espacio de medidas es un espacio compacto.

Se acaba de probar que el espacio de medidas $\mathcal{M}(X)$ es un espacio métrico. A continuación se probará que también es un espacio compacto. Para probar que el espacio de medidas es un espacio compacto, probaremos el siguiente criterio para asegurar cuándo converge una sucesión de medidas.

Proposición 12. *Sea (X, ρ) un espacio métrico compacto y $(\mathcal{M}(X), d)$ el espacio de medidas en X, donde d es la métrica definida en 1.6. Sea también $(\mu_n)_{n\in\mathbb{N}}$ una sucesión de medidas en $\mathcal{M}(X)$ y $\mu \in \mathcal{M}(X)$, entonces tenemos que $d(\mu_n, \mu) \to 0$ cuando $n \to \infty$ si y sólo si,*

$$\lim_{n\to\infty} \int \phi \mathrm{d}\mu_n = \int \phi \mathrm{d}\mu \qquad \forall \phi \in C(X).$$

Demostración. Supongamos primero que $d(\mu_n, \mu) \to 0$ si $n \to \infty$. Entonces por definición se tiene,

$$\sum_{m=1}^{\infty} \frac{1}{2^m} \left| \int \phi_m \mathrm{d}\mu_n - \int \phi_m \mathrm{d}\mu \right| \to 0 \qquad \text{si } n \to \infty \qquad \forall m \in \mathbb{N}.$$

Esto ocurre si y sólo si,

$$\left| \int \phi_m \mathrm{d}\mu_n - \int \phi_m \mathrm{d}\mu \right| \to 0 \qquad \text{si } n \to \infty \qquad \forall m \in \mathbb{N}.$$

Es decir,

$$\lim_{n\to\infty} \int \phi_m \mathrm{d}\mu_n = \int \phi_m \mathrm{d}\mu \qquad \forall m \in \mathbb{N}.$$

Consideremos $\phi \in C(X)$ no cero y $\epsilon > 0$. Como $H = (\phi_n)_{n\in\mathbb{N}}$ es denso en $\{\phi \in C(X) : ||\phi||_\infty \leq 1\}$, entonces tenemos que existe $m \in \mathbb{N}$ tal que,

$$\left\| \frac{\phi}{\|\phi\|_\infty} - \phi_m \right\|_\infty < \epsilon.$$

Entonces, para cada $n \in \mathbb{N}$ se tiene

$$\frac{1}{\|\phi\|_\infty} \left| \int \phi \mathrm{d}\mu_n - \int \phi \mathrm{d}\mu \right| \leq \left| \int \phi_m \mathrm{d}\mu_n - \int \phi_m \mathrm{d}\mu \right| + 2\epsilon.$$

Haciendo $n \to \infty$ obtenemos,

$$\frac{1}{\|\phi\|_\infty} \limsup_{n\to\infty} \left| \int \phi \mathrm{d}\mu_n - \int \phi \mathrm{d}\mu \right| \leq 2\epsilon.$$

Como ϵ es arbitrario, se concluye,

$$\lim_{n\to\infty} \int \phi \mathrm{d}\mu_n = \int \phi \mathrm{d}\mu \qquad \forall \phi \in C(X).$$

Ahora, recíprocamente supongamos que

$$\lim_{n\to\infty}\int \phi \mathrm{d}\mu_n = \int \phi \mathrm{d}\mu \qquad \forall \phi \in C(X).$$

Sea $\epsilon > 0$. Veamos que existe $k = k(\epsilon) \in \mathbb{N}$ tal que si $n > N(\epsilon)$ se tiene que $d(\mu_n, \mu) < \epsilon$. Ahora bien, por la *Propiedad Arquimediana*, sabemos que existe un número natural $m = m(\epsilon)$ tal que,

$$\frac{1}{2^{m-1}} < \epsilon.$$

Observamos que para cada $i \in \{1, \cdots, m\}$, existe $k_i \in \mathbb{N}$ tal que

$$\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| < \epsilon.$$

Tomamos $k_m^* = \max\{k_i : i \in \{1, \cdots, m\}\}$. Entonces se tiene que si $n > k_m^*$, entonces

$$\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| < \epsilon \qquad \forall i \in \{1, \cdots, m\}.$$

Así pues,

$$\begin{aligned}
d(\mu_n, \mu) &= \sum_{i=1}^{\infty} \frac{1}{2^i}\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| \\
&= \sum_{i=1}^{m} \frac{1}{2^i}\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| + \sum_{i=m+1}^{\infty} \frac{1}{2^i}\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| \\
&\leq \epsilon \sum_{i=1}^{m} \frac{1}{2^i} + \sum_{i=m+1}^{\infty} \frac{1}{2^i}\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| \\
&< \epsilon + \sum_{i=m+1}^{\infty} \frac{1}{2^i}\left|\int \phi_i \mathrm{d}\mu_n - \int \phi_i \mathrm{d}\mu\right| \\
&\leq \epsilon + \sum_{i=m+1}^{\infty} \frac{2}{2^i} = \epsilon + \frac{1}{2^{m-1}} \qquad \forall n > k_m^*.
\end{aligned}$$

La última linea se sigue de la cota encontrada en el Lema 1. Finalmente, por la elección de m, sabemos que $\frac{1}{2^{m-1}} < \epsilon$ para cada $n > k_m^*$. Así pues, se tiene que,

$$d(\mu_n, \mu) < 2\epsilon \qquad \forall n > k_m^*.$$

Por lo tanto, concluimos que $d(\mu_n, \mu) \to 0$ cuando $n \to \infty$. $\square$

El siguiente resultado muestra que el espacio $(\mathcal{M}, d)$ es un espacio compacto. La prueba del teorema ocupa el Teorema de Representación de Riesz de Teoría de la Medida, mismo que enunciamos sin prueba a continuación.

Teorema 7 (Representación de Riesz)**.** *Sea (X, ρ) un espacio métrico compacto y $J : C(X) \to \mathbb{R}$ un operador lineal, continuo y positivo. Entonces existe exactamente una medida μ en la σ-álgebra de Borel de X tal que*

$$J(\phi) = \int \phi \mathrm{d}\mu \qquad \forall \phi \in C(X).$$

Más aún, si se tiene que $J(1) = 1$, entonces $\mu(X) = 1$.

Teorema 8. *Si (X, ρ) es un espacio métrico compacto, entonces $(\mathcal{M}(X), d)$ es compacto.*

Demostración. Como $(\mathcal{M}(X), d)$ es un espacio métrico, para probar que $\mathcal{M}(X)$ es compacto, basta probar que toda sucesión $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(X)$ tiene una subsucesión convergente.

Sea $(\mu_n)_{n\in\mathbb{N}}$ una sucesión en $\mathcal{M}(X)$. Por el Teorema 6 se tiene que $C(X)$ es separable, entonces existe un conjunto $H = (\phi_n)_{n\in\mathbb{N}}$ denso en $\{\phi \in C(X) : \|\phi\|_\infty \leq 1\}$. Así pues, $\forall n \in \mathbb{N}$, definimos la sucesión $\beta_n : \mathbb{N} \to [-1, 1]$ como sigue:

$$\beta_n(k) = \int \phi_k \mathrm{d}\mu_n.$$

Por el *Teorema de Tychonoff*, se sabe que el espacio de sucesiones $[0, 1]^{\mathbb{N}}$ con la topología producto es compacto. Así pues, como $(\beta_n)_{n\in\mathbb{N}} \subseteq [0, 1]^{\mathbb{N}}$, entonces existe una subsucesión convergente $(\beta_{m_n})_{n\in\mathbb{N}}$ en esta topología. Así pues,

$$(\beta_{m_n}(k))_{n\in\mathbb{N}} = \left(\int \phi_k \mathrm{d}\mu_{m_n}\right)_{n\in\mathbb{N}}$$

converge para cada $k \in \mathbb{N}$. Así también, por densidad de H, se sabe que para cada $\phi \in C(X)$ y $\epsilon > 0$, existe $k \in \mathbb{N}$ tal que,

$$\left\|\frac{\phi}{\|\phi\|_\infty} - \phi_k\right\|_\infty < \epsilon.$$

Ahora bien, se tiene lo siguiente:

$$\begin{aligned}\left|\int \frac{\phi}{\|\phi\|_\infty}\mathrm{d}\mu_{m_n} - \int \phi_k \mathrm{d}\mu_{m_n}\right| &\leq \int \left|\frac{\phi}{\|\phi\|_\infty} - \phi_k\right| \mathrm{d}\mu_{m_n} \\ &\leq \int \left\|\frac{\phi}{\|\phi\|_\infty} - \phi_k\right\|_\infty \mathrm{d}\mu_{m_n} \\ &< \epsilon \int \mathrm{d}\mu_{m_n} \\ &= \epsilon\mu_{m_n}(X) = \epsilon.\end{aligned}$$

Así pues,

$$-\epsilon < \int \frac{\phi}{\|\phi\|_\infty}\mathrm{d}\mu_{m_n} - \int \phi_k \mathrm{d}\mu_{m_n} < \epsilon.$$

Análogamente,

$$-\epsilon < -\int \frac{\phi}{\|\phi\|_\infty} \mathrm{d}\mu_{m_l} + \int \phi_k \mathrm{d}\mu_{m_l} < \epsilon.$$

Sumamos las desigualdades,

$$\left| \int \frac{\phi}{\|\phi\|_\infty} \mathrm{d}\mu_{m_n} - \int \frac{\phi}{\|\phi\|_\infty} \mathrm{d}\mu_{m_l} - \left(-\int \phi_k \mathrm{d}\mu_{m_l} + \int \phi_k \mathrm{d}\mu_{m_n} \right) \right| < 2\epsilon.$$

De la desigualdad del triángulo $||a| - |b|| \leq |a - b|$ tenemos,

$$\left| \int \frac{\phi}{\|\phi\|_\infty} \mathrm{d}\mu_{m_n} - \int \frac{\phi}{\|\phi\|_\infty} \mathrm{d}\mu_{m_l} \right| - \left| \int \phi_k \mathrm{d}\mu_{m_n} - \int \phi_k \mathrm{d}\mu_{m_l} \right| < 2\epsilon.$$

Entonces,

$$\frac{1}{\|\phi\|_\infty} \left| \int \phi \mathrm{d}\mu_{m_n} - \int \phi \mathrm{d}\mu_{m_l} \right| < \left| \int \phi_k \mathrm{d}\mu_{m_n} - \int \phi_k \mathrm{d}\mu_{m_l} \right| + 2\epsilon \qquad \forall n, l \in \mathbb{N}.$$

Como $\left(\int \phi_k \mathrm{d}\mu_{m_n}\right)_{n\in\mathbb{N}}$ converge $\forall k \in \mathbb{N}$, entonces es de *Cauchy*. Esto implica que para m, l suficientemente grandes se tiene que,

$$\left| \int \phi_k \mathrm{d}\mu_{m_n} - \int \phi_k \mathrm{d}\mu_{m_l} \right| < \epsilon.$$

Esto implica que,

$$\frac{1}{\|\phi\|_\infty} \left| \int \phi \mathrm{d}\mu_{m_n} - \int \phi \mathrm{d}\mu_{m_l} \right| < 3\epsilon.$$

Entonces,

$$\left| \int \phi \mathrm{d}\mu_{m_n} - \int \phi \mathrm{d}\mu_{m_l} \right| < 3\epsilon \|\phi\|_\infty.$$

Por lo tanto, $\left(\int \phi \mathrm{d}\mu_{m_n}\right)_{n\in\mathbb{N}}$ es de Cauchy y como $(\mathbb{R}, |\cdot|)$ es *completo*, entonces se tiene,

$$\lim_{n\to\infty} \int \phi \mathrm{d}\mu_{m_n} \qquad \text{existe } \forall \phi \in C(X).$$

Así pues, tiene sentido definir una funcional $J : C(X) \to \mathbb{R}$ como sigue:

$$J(\phi) = \lim_{n\to\infty} \int \phi \mathrm{d}\mu_{m_n}.$$

Ahora se probará que la funcional satisface las condiciones del *Teorema de Representación de Riesz*, es decir, que es lineal, continua, positiva y que $J(1) = 1$. La funcional J es una funcional lineal por las propiedades de linealidad del límite y las propiedades de linealidad de la integral. Veamos que la funcional es continua.

Sea $\epsilon > 0$, tomamos $\delta = \epsilon$. Supongamos que $\|\phi-\psi\|_\infty < \delta$ para $\phi, \psi \in C(X)$, entonces se tiene lo siguiente:

$$\begin{aligned}
|J(\phi) - J(\psi)| &= \left| \lim_{n\to\infty} \int \phi \mathrm{d}\mu_{m_n} - \lim_{n\to\infty} \int \psi \mathrm{d}\mu_{m_n} \right| \\
&= \left| \lim_{n\to\infty} \int \phi - \psi \mathrm{d}\mu_{m_n} \right| \\
&= \lim_{n\to\infty} \left| \int \phi - \psi \mathrm{d}\mu_{m_n} \right| \\
&\leq \lim_{n\to\infty} \int |\phi - \psi| \mathrm{d}\mu_{m_n} \\
&\leq \lim_{n\to\infty} \|\phi - \psi\|_\infty \mu_{m_n}(X) \\
&= \|\phi - \psi\|_\infty .
\end{aligned}$$

Pues $\mu_{m_n}(X) = 1$. Así pues, para cada $n \in \mathbb{N}$ y como $\|\phi - \psi\|_\infty < \delta = \epsilon$ se concluye que la funcional J es continua.Veamos ahora que $J(1) = 1$.

$$J(1) = \lim_{n\to\infty} \int 1 \mathrm{d}\mu_{m_n} = \lim_{n\to\infty} \mu_{m_n}(X) = 1.$$

Supongamos que $\phi \geq 0$ y veamos que $J(\phi) \geq 0$. Por propiedades de la integral se tiene,

$$\int \phi \mathrm{d}\mu_{m_n} \geq 0 \qquad \forall n \in \mathbb{N}.$$

Así pues,

$$J(\phi) = \lim_{n\to\infty} \int \phi \mathrm{d}\mu_{m_n} \geq 0.$$

Por el Teorema de Representación de Riesz tenemos que existe una única medida μ tal que $\mu(X) = 1$ que satisface,

$$\lim_{n\to\infty} \int \phi \mathrm{d}\mu_{m_n} = J(\phi) = \int \phi \mathrm{d}\mu \qquad \forall \phi \in C(X).$$

Finalmente, por la Proposición 12, se concluye que $d(\mu_{m_n}, \mu) \to 0$ si $n \to \infty$. Por lo tanto, $\mathcal{M}(X)$ es compacto. $\square$

El espacio de medidas es un espacio convexo.

Hasta ahora se ha probado que el espacio de medidas $\mathcal{M}(X)$ es un espacio métrico compacto. En esta sección probaremos una propiedad geométrica del espacio de medidas, a saber, que el espacio $\mathcal{M}(X)$ es un espacio convexo.

Recordamos que el espacio $\mathcal{M}(X)$ es ***convexo*** si para cualesquiera dos puntos μ, ν en el espacio, se tiene que el "segmento" se recta que los une $S_{\mu\nu} = \{t\mu + (1-t)\nu : t \in [0,1]\}$ está completamente contenido en el espacio $\mathcal{M}(X)$, es decir, $S \subseteq \mathcal{M}(X)$.

Teorema 9. *Sea (X, ρ) un espacio métrico compacto. Entonces, el espacio de medidas finitas $\mathcal{M}(X)$ es un espacio convexo.*

Demostración. Sean $\mu, \nu \in \mathcal{M}(X)$. Para probar que el espacio $\mathcal{M}(X)$ es convexo, se debe verificar que $t\mu + (1-t)\nu \in \mathcal{M}(X)$ para cada $t \in [0,1]$. Sea $t \in [0,1]$, se verá que $\gamma = t\mu + (1-t)\nu$ es una medida finita de X, es decir, que satisface las propiedades para ser una medida y que la medida del espacio X es igual a 1. Así, γ satisface la primera condición, es decir, satisface que $\gamma(\emptyset) = 0$. En efecto,

$$\begin{aligned}\gamma(\emptyset) &= (t\mu + (1-t)\nu)(\emptyset) \\ &= t\mu(\emptyset) + (1-t)\nu(\emptyset) \\ &= 0.\end{aligned}$$

Sea $E \in S$. Para verificar que γ satisface la segunda condición, se debe cumplir que $\gamma(E) \geq 0$. Lo primero que se observa es que $\mu(E) \geq 0$ y $\nu(E) \geq 0$ pues μ, ν son medidas. Así también, dado que $t \in [0,1]$, se tiene que $t \geq 0$ y que $1 - t \geq 0$. Entonces,

$$t\mu(E) \geq 0.$$

Además,

$$(1-t)\nu(E) \geq 0.$$

Esto implica que,

$$\gamma(E) = (t\mu + (1-t)\nu)(E) \geq 0.$$

Sea $(E_n)_{n\in\mathbb{N}} \subseteq S$ es una sucesión de elementos ajenos. Para ver que γ satisface la tercera condición, se debe verificar que se cumple que $\gamma(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \gamma(E_n)$. En efecto,

$$\begin{aligned}\gamma\left(\bigcup_{n=1}^{\infty} E_n\right) &= (t\mu + (1-t)\nu)\left(\bigcup_{n=1}^{\infty} E_n\right) \\ &= t\mu\left(\bigcup_{n=1}^{\infty} E_n\right) + (1-t)\nu\left(\bigcup_{n=1}^{\infty} E_n\right) \\ &= t\sum_{n=1}^{\infty} \mu(E_n) + (1-t)\sum_{n=1}^{\infty} \nu(E_n) \\ &= \sum_{n=1}^{\infty} (t\mu + (1-t)\nu)(E_n) \\ &= \sum_{n=1}^{\infty} \gamma(E_n).\end{aligned}$$

Concluimos que γ es una medida. Finalmente, hay que verificar que se cumple $\gamma(X) = 1$. Como $\mu, \nu \in \mathcal{M}(X)$, entonces se tiene que $\mu(X) = 1 = \nu(X)$. Así pues,

$$\begin{aligned}\gamma(X) &= (t\mu + (1-t)\nu)(X) \\ &= t\mu(X) + (1-t)\nu(X) \\ &= t + 1 - t = 1.\end{aligned}$$

Por lo tanto, $\gamma \in \mathcal{M}(X)$ y entonces se tiene que $\mathcal{M}(X)$ es un espacio convexo.

$\square$

1.3.2. Teorema de Kyrilov-Bogolubov

El Teorema de Kyrilov-Bogolubov resuelve el problema de las existencia de medidas invariantes. El teorema prueba que dada una transformación continua T, existe una medida μ tal que T preserva la medida μ, siempre y cuando el espacio X sea un espacio métrico compacto. Antes de probar el teorema, probaremos algunos resultados auxiliares. El primer resultado que probaremos nos afirma que la función característica de un conjunto compuesta con una transformación T es igual a la característica de la imagen inversa del conjunto.

Lema 2. *Sea X un conjunto, $B \subseteq X$ y $T : X \to X$ una transformación. Entonces,*

$$\chi_B \circ T = \chi_{T^{-1}(B)}.$$

Demostración.

$$\begin{aligned}
\chi_B \circ T(x) &= \chi_B(T(x) \\
&= \begin{cases} 1, \text{si } T(x) \in B \\ 0, \text{si } T(x) \notin B \end{cases} \\
&= \begin{cases} 1, \text{si } x \in T^{-1}(B) \\ 0, \text{si } x \notin T^{-1}(B) \end{cases} \\
&= \chi_{T^{-1}(B)}(x).
\end{aligned}$$

$\square$

Ahora, vamos a considerar una transformación $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ del espacio de medidas finitas en sí mismo, definida de la siguiente manera,

$$\begin{aligned}
\mu &\mapsto \mu \circ T^{-1} \\
(T_*\mu)(A) &= \mu(T^{-1}(A)).
\end{aligned} \tag{1.9}$$

Veamos que la transformación T_* satisface la propiedad de que para cada función, la integral con respecto a $T_*\mu$ coincide con la integral de la función compuesta con T respecto a μ, es decir, Si $\phi \in L^1(X, \mu)$, entonces,

$$\int \phi \mathrm{d}(T_*\mu) = \int \phi \circ T \mathrm{d}\mu.$$

Para probar esta propiedad ocuparemos dos resultados centrales de la Teoría de la Medida. El primer resultado se conoce como el *Lema Básico de Aproximación* y el segundo es el *Teorema de la Convergencia Monótona.* A continuación enunciamos sin prueba ambos resultados.

Teorema 10 (Lema Básico de Aproximación)**.** *Sea (X, S) un espacio medible y $f : X \to \mathbb{R}$ una función medible no negativa, entonces existe una sucesión $(s_n)_{n\in\mathbb{N}}$ de funciones simples tales que:*

- $0 \leq s_n \leq s_{n+1} \leq f$.
- $f(x) = \lim_{n\to\infty} s_n(x)$ *para cada* $x \in X$.
- *Si* f *es acotada, entonces* $s_n \to f$ *de manera uniforme en* X.

Teorema 11 (de la Convergencia Monótona)**.** *Sea* (X, S, μ) *un espacio de medida y* $(f_n)_{n\in\mathbb{N}}$ *una sucesión no decreciente de funciones medibles tal que converge a la función medible* f*. Entonces se cumple la siguiente propiedad.*

$$\int_E f \mathrm{d}\mu = \int_E \lim_{n\to\infty} f_n \mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \mathrm{d}\mu \qquad \forall E \in S.$$

Probaremos el siguiente lema que se ocupará en la prueba del Teorema de Kyrilov-Bogolubov.

Lema 3. *Sea* (X, d) *un espacio métrico compacto,* S *una* σ*-álgebra y* $T\colon X \to X$ *una transformación continua. Si* $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ *se define como en (1.9), entonces se cumple lo siguiente:*

$$\int \phi \mathrm{d}(T_*\mu) = \int \phi \circ T \mathrm{d}\mu \qquad \forall \phi \in L^1(X, \mu).$$

Demostración. Sea $B \subseteq X$ un conjunto medible. Consideremos la función característica χ_B. Tenemos,

$$\int \chi_B \mathrm{d}(T_*\mu) = (T_*\mu)(B) = \mu(T^{-1}(B)) = \int \chi_{T^{-1}(B)} \mathrm{d}\mu.$$

Pero se probó anteriormente que $\chi_B \circ T = \chi_{T^{-1}(B)}$. Por lo tanto, se tiene,

$$\int \chi_B \mathrm{d}(T_*\mu) = \int \chi_B \circ T \mathrm{d}\mu.$$

Consideremos una función simple $s : X \to \mathbb{R}$. Las funciones simples se ven de la forma $\sum_{i=1}^n \alpha_i \chi_{E_i}$ con $\alpha_i \in \mathbb{R}$, E_i en el álgebra de conjuntos $\forall i \in \{1, \cdots, n\}$ y $n \in \mathbb{N}$. Entonces tenemos lo siguiente:

$$\begin{aligned}
\int s \mathrm{d}(T_*\mu) &= \int \sum_{i=1}^n \alpha_i \chi_{E_i} \mathrm{d}(T_*\mu) \\
&= \sum_{i=1}^n \alpha_i \int \chi_{E_i} \mathrm{d}(T_*\mu) \\
&= \sum_{i=1}^n \alpha_i \int \chi_{E_i} \circ T \mathrm{d}\mu \\
&= \int \sum_{i=1}^n \alpha_i (\chi_{E_i} \circ T) \mathrm{d}\mu \\
&= \int s \circ T \mathrm{d}\mu.
\end{aligned}$$

Sea $f \in M(X,\mu)$. Por el Lema Básico de Aproximación, se sabe que existe una sucesión de funciones simples $(s_n)_{n\in\mathbb{N}}$ tales que $s_n \to f$ cuando $n \to \infty$. Así también, por el Teorema de la Convergencia Monótona,

$$\int s_n \mathrm{d}(T_*\mu) \to \int \phi \mathrm{d}(T_*\mu)$$

y

$$\int s_n \circ T \mathrm{d}\mu \to \int f \circ T \mathrm{d}\mu,$$

ya que $\int s_n \mathrm{d}(T_*\mu) = \int s_n \circ T\mathrm{d}\mu$. Por la unicidad de los límites, tenemos que,

$$\int f\mathrm{d}(T_*\mu) = \int (f\circ T)\mathrm{d}\mu.$$

Sea $\phi \in L^1(X, T_*\mu)$. Sabemos que ϕ se puede escribir como $\phi = \phi^+ - \phi^-$, con $\phi^+, \phi^- \in M(X,S)$. Tenemos lo siguiente,

$$\begin{aligned}
\int \phi \mathrm{d}(T_*\mu) &= \int \phi^+ \mathrm{d}(T_*\mu) - \int \phi^- \mathrm{d}(T_*\mu) \\
&= \int \phi^+ \circ T\mathrm{d}(T_*\mu) - \int \phi^- \circ T\mathrm{d}(T_*\mu) \\
&= \int (\phi^+ - \phi^-) \circ T\mathrm{d}(T_*\mu) \\
&= \int (\phi\circ T)\mathrm{d}\mu.
\end{aligned}$$

□

La prueba del Teorema de Kyrilov-Bogolubov consiste en encontrar los puntos fijos de la transformación T_*. Probaremos que la transformación T_* siempre tiene puntos fijos y así, siempre hay medidas invariantes.

Teorema 12 (Kyrilov-Bogolubov)**.** *Sea (X,d) un espacio métrico compacto y S una σ-álgebra. Si $T : X \to X$ es una transformación continua, entonces existe una medida $\mu \in \mathcal{M}(X)$ tal que T preserva la medida μ.*

Demostración. Consideramos la transformación $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ definida como:

$$\mu \mapsto \mu \circ T^{-1}$$
$$(T_*\mu)(A) = \mu(T^{-1}(A)).$$

Observamos que si la transformación T_* tiene un punto fijo μ, entonces μ es T-invariante. Así pues, probaremos que T_* tiene un punto fijo. Veamos que T_* es continua. Por el lema anterior, se sabe que

$$\int \phi \mathrm{d}(T_*\mu) = \int \phi \circ T\mathrm{d}\mu \qquad \forall \phi \in L^1(X,\mu).$$

Como $C(X) \subseteq L^1(X,\mu)$ se tiene que

$$\int \phi \mathrm{d}(T_*\mu) = \int (\phi \circ T)\mathrm{d}\mu \qquad \forall \phi \in C(X).$$

Por hipótesis X es un espacio métrico compacto, entonces, por el Teorema 8, se tiene que $\mathcal{M}(X)$ es compacto y por ende, es cerrado, asi pues, para cada $\mu \in \mathcal{M}(X)$ sabemos que existe una sucesión $(\mu_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}(X)$ tal que $\mu_n \to \mu \in \mathcal{M}(X)$ y por la Proposición 12 se tiene que

$$\int \phi \mathrm{d}\mu_n \to \int (\phi \circ T)\mathrm{d}\mu \qquad \text{si } n \to \infty \ \forall \phi \in C(X). \tag{1.10}$$

Por hipótesis, sabemos que T es continua. Esto implica que $\phi \circ T$ es continua. Por (1.10) tenemos que,

$$\int (\phi \circ T)\mathrm{d}\mu_n \to \int (\phi \circ T)\mathrm{d}\mu \qquad \text{si } n \to \infty.$$

Como $\int (\phi \circ T)\mathrm{d}\mu = \int \phi \mathrm{d}(T_*\mu)$, por la Proposición 12 concluimos que

$$T_*\mu_n \to T_*\mu \qquad \text{si } n \to \infty.$$

Por lo tanto, T_* es continua. Ahora, sea $\mu \in \mathcal{M}(X)$. Para cada $n \in \mathbb{N}$, definimos la sucesión

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \mu.$$

Veamos que μ_n es una medida para cada $n \in \mathbb{N}$. Debemos verificar que se cumple lo siguiente:

- $\mu_n(\emptyset) = 0$.
- $\mu_n(E) \geq 0 \ \forall E \in S$.
- Si $(E_l)_{l\in\mathbb{N}}$ es una sucesión de elementos disjuntos de S, entonces,

$$\mu_n\left(\bigcup_{l=1}^{\infty} E_l\right) = \sum_{l=1}^{\infty} \mu_n(E_l).$$

Así pues,

$$\begin{aligned}
\mu_n(\emptyset) &= \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \mu(\emptyset) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(\emptyset)) \\
&= \frac{1}{n}\sum_{k=0}^{n-1} \mu(\emptyset) \\
&= 0.
\end{aligned}$$

Sea $E \in S$, veamos que $\mu_n(E) \geq 0$.

$$\begin{aligned}\mu_n(E) &= \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \mu(E)\\ &= \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(E)).\end{aligned}$$

Como μ es una medida, se tiene que $\mu(T^{-k}(E)) \geq 0$ para toda k. Por lo tanto $\mu_n(E) \geq 0$. Ahora, sea $(E_l)_{l\in\mathbb{N}}$ es una sucesión de elementos ajenos de S.

$$\begin{aligned}\mu_n\left(\bigcup_{l=1}^{\infty} E_l\right) &= \frac{1}{n}\sum_{k=0}^{n-1} T_*^k \mu\left(\bigcup_{l=1}^{\infty} E_l\right)\\ &= \frac{1}{n}\sum_{k=0}^{n-1} \mu\left(T^{-k}\left(\bigcup_{l=1}^{\infty} E_l\right)\right)\\ &= \frac{1}{n}\sum_{k=0}^{n-1} \mu\left(\bigcup_{l=1}^{\infty} T^{-k}(E_l)\right)\\ &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{l=1}^{\infty} \mu(T^{-k}(E_l))\\ &= \sum_{l=1}^{\infty}\frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(E_l))\\ &= \sum_{l=1}^{\infty} \mu_n(E_l).\end{aligned}$$

Por lo tanto, μ_n es una medida para cada $n \in \mathbb{N}$. Por el Teorema 8, sabemos que $\mathcal{M}(X)$ es compacto. Así pues, existe una subsucesión $(\mu_{m_n})_{n\in\mathbb{N}}$ contenida en $(\mu_n)_{n\in\mathbb{N}}$ convergente. Supongamos que $(\mu_{m_n})_{n\in\mathbb{N}}$ converge a $\nu \in \mathcal{M}(X)$.

$$
\begin{aligned}
T_*\mu_{m_n} &= T_*\left(\frac{1}{m_n}\sum_{k=0}^{m_n-1} T_*^k\mu\right)\\
&= \frac{1}{m_n}\sum_{k=0}^{m_n-1} T_*^k\mu \circ T^{-1}\\
&= \frac{1}{m_n}\sum_{k=0}^{m_n-1} \mu \circ T^{-(k+1)}\\
&= \frac{1}{m_n}\sum_{k=1}^{m_n} T_*^k\mu\\
&= \frac{T_*\mu}{m_n} + \frac{T_*^2\mu}{m_n} + \cdots + \frac{T_*^{m_n}\mu}{m_n}\\
&= \frac{T_*\mu}{m_n} + \frac{T_*^2\mu}{m_n} + \cdots + \frac{T_*^{m_n}\mu}{m_n} + \frac{\mu}{m_n} - \frac{\mu}{m_n}\\
&= \frac{\mu}{m_n} + \frac{T_*\mu}{m_n} + \frac{T_*^2\mu}{m_n} + \cdots + \frac{T_*^{m_n}\mu}{m_n} - \frac{\mu}{m_n}\\
&= \mu_{m_n} + \frac{T_*^{m_n}\mu}{m_n} - \frac{\mu}{m_n}.
\end{aligned}
$$

Si $n \to \infty$, entonces $\frac{T_*^{m_n}\mu}{m_n} \to 0$, $\frac{\mu}{m_n} \to 0$ y $\mu_{m_n} \to \nu$. Por lo tanto,

$$T_*\mu_{m_n} \to \nu \qquad \text{si } n \to \infty.$$

Como T_* es continua, entonces,

$$T_*\mu_{m_n} \to T_*\nu \qquad \text{si } n \to \infty.$$

Por unicidad de los límites se tiene que $T_*\nu = \nu$. Por lo tanto, ν es un punto fijo y concluimos que existe una medida T-invariante en X. $\square$

Capítulo 2

Teoremas Ergódicos

En este capítulo se prueban los teoremas más importantes de la Teoría Ergódica. Primero se prueba el *Teorema de la Recurrencia de Poincaré*, y posteriormente se prueba el *Teorema Ergódico de Birkhoff*. Asimismo se estudian algunos ejemplos donde se aprecia la aplicación de dichos teoremas.

2.1. Teorema de Recurrencia de Poincaré

En el estudio de los sistemas dinámicos se busca conocer el comportamiento de los sistemas a través del tiempo. Se suele representar la evolución a lo largo del tiempo de de cada punto de un espacio X de una manera discreta. Consideramos cada iteración de una transformación $T : X \to X$ como una unidad de tiempo. Por ejemplo, si consideramos que x_0 es un punto en el tiempo 0, entonces $x_1 = T(x_0)$ es el mismo punto después de una unidad de tiempo y $x_2 = T(x_1) = T^2(x_0)$ es el punto después de dos unidades de tiempo. De manera general $x_n = T(x_{n-1}) = T^n(x_0)$ es el punto después de que han pasado n unidades de tiempo. A esta serie de iteraciones $(T^n(x_0))_{n\in\mathbb{N}}$ de un punto inicial x_0 se le conoce como la *órbita* de la condición inicial x_0.

El Teorema de la Recurrencia de Poincaré afirma que las órbitas de los puntos de los conjuntos medibles de un espacio de medida X son *recurrentes*. Al iterar los puntos de un conjunto medible E bajo la acción de una transformación T, para cada punto $x \in E$ existirá una $n \in \mathbb{N}$ tal que $T^n(x) \in E$ (salvo los puntos de un conjunto de medida cero). Más aún, no sólo se puede asegurar que la órbita de cada punto en E regresará al menos una vez al conjunto original E, sino que de hecho hay un número infinito de valores de n tales que $T^n(x) \in E$. Una manera de expresarlo es diciendo que si consideráramos la serie

$$\sum_{n=0}^{\infty} \chi_E(T^n(x)),$$

donde χ_E es la función característica de E, encontraríamos que la serie diverge para cada punto en E, salvo un conjunto de medida cero. A continuación, se enuncia y prueba de manera formal el Teorema de la Recurrencia de Poincaré.

Teorema 13 (Recurrencia de Poincaré)**.** *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$ y sea $T : X \to X$ una transformación medible que preserva la medida μ. Si $E \subset X$ es un conjunto medible, entonces el conjunto*

$$B = \{x \in E : T^n(x) \in E \textit{ para una cantidad infinita de } n \in \mathbb{N}\}$$

tiene la misma medida que E, es decir, $\mu(B) = \mu(E)$.

Demostración. Por hipótesis, tenemos que

$$B = \{x \in E : T^n(x) \in E \text{ para una cantidad infinita de } n \in \mathbb{N}\}.$$

Así pues,

$$\begin{aligned}
B &= E \cap \limsup_{n\to\infty} T^{-n}(E) \\
&= E \cap \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(E) \\
&= (E \cap (X \setminus E)) \cup \left(E \cap \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(E)\right) \\
&= E \cap \left((X \setminus E) \cup \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(E)\right) \\
&= E \cap \left((X \setminus E) \cup X \setminus \left(X \setminus \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right)\right) \\
&= E \cap \left(X \setminus \left(E \cap \left(X \setminus \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right)\right) \\
&= E \setminus E \cap \left(X \setminus \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}(E)\right) \\
&= E \setminus \left(E \cap \bigcup_{n=1}^{\infty} \left(X \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right) \\
&= E \setminus \left(\bigcup_{n=1}^{\infty} \left(E \cap \left(X \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right)\right) \\
&= E \setminus \left(\bigcup_{n=1}^{\infty} \left(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right).
\end{aligned}$$

Como por hipótesis $\mu(X) = 1 < \infty$, se tiene lo siguiente

$$\begin{aligned}
\mu(B) &= \mu(E) - \mu\left(\bigcup_{n=1}^{\infty} \left(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right) \\
&\leq \mu(E).
\end{aligned}$$

Por otra parte, se tiene también que como

$$\mu\left(\bigcup_{n=1}^{\infty}\left(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right)\right) \leq \sum_{n=1}^{\infty} \mu\left(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right),$$

entonces,

$$\mu(E) - \sum_{n=1}^{\infty} \mu\left(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right) \leq \mu(B).$$

Para probar que $\mu(E) \leq \mu(B)$, veamos que $\mu(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)) = 0$ para cada $n \in \mathbb{N}$. Observamos que

$$E \subseteq \bigcup_{k=0}^{\infty} T^{-k}(E).$$

Entonces

$$E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E) \subseteq \bigcup_{k=0}^{\infty} T^{-k}(E) \setminus \bigcup_{k=n}^{\infty} T^{-k}(E).$$

Afirmación: $T^{-n}(\bigcup_{k=0}^{\infty} T^{-k}(E)) = \bigcup_{k=n}^{\infty} T^{-k}(E)$.

$$x \in T^{-n}\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right) \text{ si y sólo si } T^{n}(x) \in \left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right).$$

Así pues, existe $k' \in \mathbb{N}$ tal que,

$$T^{n}(x) \in T^{-k'}(E) \text{ si y sólo si } x \in T^{-(n+k')}(E),$$

lo cual a su vez está contenido en $\bigcup_{k=n}^{\infty} T^{-k}(E)$. Reescribimos la ecuación anterior,

$$E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E) \subseteq \bigcup_{k=0}^{\infty} T^{-k}(E) \Bigg\backslash T^{-n}\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right)$$

Esto implica que,

$$\begin{aligned}\mu\left(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)\right) &\leq \mu\left(\bigcup_{k=0}^{\infty} T^{-k}(E) \Bigg\backslash T^{-n}\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right)\right)\\ &= \mu\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right) - \mu\left(T^{-n}\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right)\right).\end{aligned}$$

Por hipótesis, μ es T-invariante,

$$\mu\left(T^{-n}\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right)\right) = \mu\left(\bigcup_{k=0}^{\infty} T^{-k}(E)\right).$$

Por lo tanto $\mu(E \setminus \bigcup_{k=n}^{\infty} T^{-k}(E)) = 0$ y $\mu(B) = \mu(E)$. □

La primera observación que hacemos respecto al Teorema de la Recurrencia de Poincaré es que la hipótesis de que la medida del espacio debe ser finita es una hipótesis imprescindible. A continuación se muestra un ejemplo donde la medida del espacio es infinita y no se cumple el Teorema.

Ejemplo 7. Consideremos el espacio de medida $(\mathbb{R}, S, \lambda)$, donde S es la σ-álgebra de Borel y λ es la medida de Lebesgue. Sabemos que $\lambda(\mathbb{R}) = \infty$, es decir, no es un espacio de medida finita. Consideramos la transformación $T : X \to X$ definida como sigue,

$$T(x) = x + 1.$$

Consideramos el conjunto $E = (0, 1)$. Sabemos que $E \in S$ y más aun, que $\lambda(E) = 1$. Sin embargo, al aplicar la transformación T al conjunto E se tiene,

$$\begin{aligned} E &= (0,1) \\ T(E) &= (1,2) \\ &\vdots \\ T^n(E) &= (n, n+1) \\ &\vdots \end{aligned}$$

Así pues, $T^n(E) \cap T^m(E) = \emptyset$ para $n \neq m$. Por lo tanto, el conjunto

$$B = \{x \in E : T^n(x) \in E \textit{ para una cantidad infinita de } n \in \mathbb{N}\} = \emptyset$$

Entonces $\mu(B) = \mu(\emptyset) = 0$.

El siguiente ejemplo presenta un espacio compacto en el que la medida del espacio tampoco es finita y por ende, no se cumple el Teorema de la Recurrencia de Poincaré.

Ejemplo 8. Consideremos el espacio de medida $([0, 1]/\{0, 1\}, S, \mu)$, donde S es la σ-álgebra de Borel y μ es la medida del conteo. Así pues, $\mu([0, 1]/\{0, 1\}) = \infty$, es decir, no es un espacio de medida finita. Sea $\alpha \in (0, 1) \cap \mathbb{I}$. Consideramos la transformación traslación,

$$T_\alpha(x) = x + \alpha \mod 1.$$

Observamos que $T_\alpha^n(x) = x + n\alpha \mod 1 = T_\alpha^n(0) + x$. Entonces, $T_\alpha^n(x) \neq T_\alpha^m(x)$ para toda $n, m \in \mathbb{N}$ tales que $n \neq m$. En efecto, si $T_\alpha^n(x) = T_\alpha^m(x)$, entonces se tendría que $n\alpha \mod 1 = m\alpha \mod 1$. Como α es irracional, $n = m$. Por lo tanto, se tiene que cuando la traslación es irracional, los valores en la órbita de un punto x no se repiten.

Sea $A = \{x\}$. Como μ es la medida del conteo, se tiene que $\mu(A) = 1$, pero como los valores en la órbita no se repiten, entonces,

$$B = \{x \in A : T^n(x) \in A \textit{ para un número infinito de } n \in \mathbb{N}\} = \emptyset.$$

Por lo tanto $\mu(B) = 0 \neq 1 = \mu(A)$.

Para ilustrar la utilidad del Teorema de la Recurrencia de Poincaré, veremos una aplicación para las expansiones binarias de los números en el intervalo $(0,1)$. Probaremos que en la expansión binaria de cada número debe aparecer un número infinito de 0's y 1's.

Ejemplo 9. En la sección 1.2.1 se presentó la transformación expansión binaria $T:[0,1)\to[0,1)$ definida como $T(x)=2x \mod 1$. Veamos primero cómo es que se puede escribir la expansión binaria de un número con ayuda de la transformación T. Posteriormente con base en el Teorema de la Recurrencia de Poincaré, probaremos que la expansión binaria para casi todo número en $(0,1)$ contiene una cantidad infinita de 0's y 1's. Definimos la función $a_1:(0,1)\to\{0,1\}$ como

$$a_1(x)=\begin{cases} 0 & si \quad 0\leq x<\frac{1}{2}\\ 1 & si \quad \frac{1}{2}\leq x<1\end{cases}$$

La función $T(x)=2x \mod 1$ se puede escribir como

$$T(x)=2x-a_1(x).$$

Ahora definimos

$$a_n(x)=a_1(T^{n-1}(x)) \qquad \forall n\geq 1.$$

Sea $x\in X$ fijo (por simplicidad se escribirá a_n en lugar de $a_n(x)$). Como,

$$T(x)=2x-a_1 \qquad \Rightarrow \qquad x=\frac{a_1}{2}+\frac{T(x)}{2}.$$

Entonces,

$$\begin{aligned} T^2(x)=2T(x)-a_2 \qquad &\Rightarrow \qquad T(x)=\frac{a_2}{2}+\frac{T^2(x)}{2}\\ &\Rightarrow \qquad x=\frac{a_1}{2}+\frac{a_2}{2^2}+\frac{T^2(x)}{2^2}.\end{aligned}$$

Continuando de esta manera, se tiene que

$$x=\frac{a_1}{2}+\frac{a_2}{2^2}+\cdots+\frac{a_n}{2^n}+\frac{T^n(x)}{2^n} \qquad \forall n\geq 1.$$

Esto implica,

$$x-\sum_{i=1}^{n}\frac{a_i}{2^n}=\frac{T^n(x)}{2^n} \qquad \forall n\geq 1.$$

Como $0<T^n(x)<1$ para cada $n\geq 1$. De esto se sigue que,

$$\frac{T^n(x)}{2^n}<\frac{1}{2^n}\to 0 \qquad si\ n\to\infty.$$

Por lo tanto, $x=\sum_{i=1}^{\infty}\frac{a_i}{2^i}$. Por un argumento análogo, se tiene que,

$$T^m(x)=\sum_{i=1}^{\infty}\frac{a_{m+i}}{2^i}.$$

Para cada $n \in \mathbb{N}$ e $(i_1, \cdots, i_n) \in \{0,1\}^n$, consideramos el intervalo,

$$I_{i_1,\cdots,i_n} = \left(\sum_{j=1}^{n} \frac{i_j}{2^j}, \sum_{j=1}^{n} \frac{i_j}{2^j} + \frac{1}{2^n} \right) \subseteq [0,1].$$

Por el Teorema de la Recurrencia de Poincaré, se tiene que para casi todo $x \in I_{i_1,\cdots,i_n}$,

$$|\{m \in \mathbb{N} : T^m(x) \in I_{i_1,\cdots,i_n}\}| = \infty.$$

Como $x = \sum_{i=1}^{\infty} \frac{i}{2^i}$, $T^m(x) = \sum_{i=1}^{\infty} \frac{i_{m+1}}{2^i}$, concluimos que para casi todo $x \in I_{i_1,\cdots,i_n}$ los dígitos $i_1, \cdots, i_n$ aparecen infinitas veces en la expansión binaria de x en ese mismo orden.

2.2. Teorema Ergódico de Birkhoff

El Teorema de Recurrencia de Poincaré afirma que cuando la medida del espacio X es finita, la órbita de casi todos los puntos en un conjunto medible $E \subseteq X$ regresa una cantidad infinita de veces al conjunto E. Sin embargo, nos gustaría saber cuánto tiempo se mantienen las imágenes de los puntos recurrentes en el conjunto E. El *Teorema Ergódico de Birkhoff* nos da dicha respuesta, pues nos permite calcular la *frecuencia* con la que permanece la órbita de un punto en un conjunto $E \subseteq X$ medible.

En la sección anterior 2.1 se mencionó que por el Teorema de la Recurrencia de Poincaré, se podía asegurar que la serie

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \chi_E(T^k(x))$$

siempre divergía. Sin embargo, el Teorema Ergódico de Birkhoff asegura que la sucesión de promedios, o *sumas de Cèsaro*, convergen, es decir, la siguiente serie es convergente.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_E(T^k(x)) < \infty.$$

Una manera de entender el Teorema Ergódico de Birkhoff es verlo como una generalización del la *Ley Fuerte de los Números Grandes* de Probabilidad. Dajani [9] afirma que si consideramos una sucesión de *variables aleatorias* $X_1, X_2, \cdots$ entonces,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} X_i = E(X_1),$$

donde $E(X_1)$ es la esperanza de la primera variable.

Dependiendo el problema podemos definir una transformación medible por lo que la expresión más general consiste en calcular el límite

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)),$$

que es precisamente lo que resuelve el Teorema Ergódico de Birkhoff. En esta sección se presentará la prueba del Teorema Ergódico de Birkhoff. La prueba que se presenta se debe a A. Katok y B. Hasselblatt publicada en "Introduction to the Modern Theory of Dynamical Systems. Encyclopedia of Mathematics and Its Applications," y se sigue la reconstrucción hecha por L. Barreira [2].

A continuación se prueba un criterio en términos de funciones integrables para afirmar que una transformación preserva la medida. Posteriormente se definen los *conjuntos invariantes* y las *funciones invariantes*. Un tipo especial de función invariante recibe el nombre de *esperanza condicional* y se ocupará en la prueba del Teorema Ergódico de Birkhoff.

2.2.1. Criterio de Transformaciones que Preservan la Medida

Para probar el Teorema Ergódico de Birkhoff probaremos antes un criterio para afirmar que una transformación T preserva la medida que ocuparemos más adelante. La prueba del criterio ocupa el Lema Básico de Aproximación, el Teorema de la Convergencia Monótona y el Lema de Fatou. Los primeros dos resultados fueron enunciados anteriormente como el Teorema 10 y el Teorema 11. A continuación enunciamos sin prueba el Lema de Fatou.

Teorema 14 (Lema de Fatou)**.** *Sea (X, S, μ) un espacio de medida. Sea (f_n) una sucesión de funciones medibles, entonces,*

$$\int_E \liminf_{n\to\infty} f_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_E f_n \mathrm{d}\mu \qquad \forall E \in S.$$

El criterio que probaremos a continuación afirma que toda función integrable $\phi : X \to \mathbb{R}$ tal que que la composición con la transformación $\phi \circ T$ es a su vez una función integrable y su integral coincida con la de la función ϕ es una condición necesaria y suficiente para que afirmar que la transformación T preserva la medida μ.

Teorema 15 (Criterio)**.** *Sea (X, S, μ) un espacio de medida tal que $\mu(X) < +\infty$ y sea $T : X \to X$ una transformación medible. T preserva la medida μ si y sólo si para cada función $\phi \in L^1(X, \mu)$, se tiene que $\phi \circ T \in L^1(X, \mu)$ y*

$$\int \phi \circ T d\mu = \int \phi d\mu.$$

Demostración. Primero veamos que T preserva la medida μ. Sea $B \subseteq X$ un conjunto medible. Probaremos que $\mu(T^{-1}(B)) = \mu(B)$. Por hipótesis sabemos que para cada $\phi \in L^1(X, \mu)$ se cumple,

- $\phi \circ T \in L^1(X, \mu)$.
- $\int \phi \circ T d\mu = \int \phi d\mu$.

En particular, como $\chi_B \in L^1(X, \mu)$, se satisface,

- $\chi_B \circ T \in L^1(X, \mu)$.
- $\int \chi_B \circ T d\mu = \int \chi_B d\mu = \mu(B)$.

Por el Lema 2 se tiene que $\chi_B \circ T = \chi_{T^{-1}(B)}$, entonces,

$$\int \chi_B \circ T d\mu = \int \chi_{T^{-1}(B)} \mathrm{d}\mu = \mu(T^{-1}(B)).$$

Así pues, $\mu(T^{-1}(B)) = \mu(B)$. Por lo tanto, T preserva la medida μ. Recíprocamente supongamos que T preserva la medida μ. Entonces se tiene lo siguiente,

$$\int \chi_B \circ T \mathrm{d}\mu = \int \chi_B \mathrm{d}\mu \qquad \forall B \subseteq X \quad \text{conjunto medible.}$$

Sea $s : X \to \mathbb{R}$ una función simple, es decir, una función de la forma $\sum_{i=1}^{n} \alpha_i \chi_{E_i}$ con $\alpha_i \in \mathbb{R}$ y E_i en el álgebra de conjuntos para cada $i \in \{1, \cdots, n\}$ y $n \in \mathbb{N}$. Entonces,

$$\begin{aligned} \int s \circ T \mathrm{d}\mu &= \int \left(\sum_{i=1}^{n} \alpha_i \chi_{E_i} \right) \circ T \mathrm{d}\mu \\ &= \sum_{i=1}^{n} \int \alpha_i \chi_{E_i} \circ T \mathrm{d}\mu \\ &= \sum_{i=1}^{n} \alpha_i \int \chi_{E_i} \circ T \mathrm{d}\mu \\ &= \sum_{i=1}^{n} \alpha_i \int \chi_{E_i} \mathrm{d}\mu \\ &= \int s \mathrm{d}\mu. \end{aligned}$$

Por lo tanto $\int s \circ T \mathrm{d}\mu = \int s \mathrm{d}\mu$. Ahora, sea $f \in M(X, S)$. Por el Lema Básico de Aproximación sabemos que hay una sucesión de funciones simples $(s_n)_{n \in \mathbb{N}}$ creciente tales que $s_n \to f$ cuando $n \to +\infty$. Por el Teorema de la Convergencia Monótona,

$$\int s_n \mathrm{d}\mu \to \int \phi \mathrm{d}\mu \text{ cuando } n \to +\infty.$$

Así también, $s_n \circ T \nearrow f \circ T$ cuando $n \to +\infty$. Por el Lema de Fatou se tiene,

$$\begin{aligned} \int f \circ T \mathrm{d}\mu &= \int \liminf_{n \to \infty} s_n \circ T \mathrm{d}\mu \\ &\leq \liminf_{n \to \infty} \int s_n \circ T \mathrm{d}\mu \\ &= \liminf_{n \to \infty} \int s_n \mathrm{d}\mu \\ &= \int f \mathrm{d}\mu < +\infty. \end{aligned}$$

Por lo tanto, $f \circ T \in L^1(X, \mu)$. Como $s_n \circ T \nearrow f \circ T$ cuando $n \to +\infty$, por el Teorema de la Convergencia Monótona,

$$\begin{aligned}\int \lim_{n\to\infty} s_n \circ T \mathrm{d}\mu &= \int f \circ T \mathrm{d}\mu \\ &= \lim_{n\to\infty} \int s_n \circ T \mathrm{d}\mu \\ &= \lim_{n\to\infty} \int s_n \mathrm{d}\mu \\ &= \int \lim_{n\to\infty} s_n \mathrm{d}\mu \\ &= \int f \mathrm{d}\mu.\end{aligned}$$

Entonces, $\int f \circ T \mathrm{d}\mu = \int f \mathrm{d}\mu$. Ahora, sea $\phi \in L^1(X, \mu)$. Sabemos que ϕ se puede ver como $\phi = \phi^+ - \phi^-$ con $\phi^+, \phi^- \in M(X, \mu)$. Entonces,

$$\begin{aligned}\int \phi \mathrm{d}\mu &= \int \phi^+ \mathrm{d}\mu - \int \phi^- \mathrm{d}\mu \\ &= \int \phi^+ \circ T \mathrm{d}\mu - \int \phi^- \circ T \mathrm{d}\mu \\ &= \int \phi \circ T \mathrm{d}\mu\end{aligned}$$

□

2.2.2. Conjuntos Invariantes y Funciones Invariantes

En esta sección vamos a estudiar la noción de invarianza. Consideramos una transformación medible T. Por una parte, se requiere que al tomar la imagen inversa de un conjunto se obtenga el mismo conjunto. Por otra parte, se requiere que al aplicar una función ϕ a un punto x de espacio, se obtenga lo mismo que al aplicar primero la transformación a x y posteriormente aplicar la función ϕ. A continuación formalizamos estas ideas.

Definición 18. *Sea (X, S, μ) un espacio de medida finita, $T : X \to X$ una transformación medible. Definimos lo siguiente:*

- *Un conjunto $A \subseteq X$ es un **conjunto T-invariante** si y sólo si $T^{-1}(A) = A$.*
- *Una función $\phi : X \to \mathbb{R}$ es una **función T-invariante** si y sólo si $\phi(T(x)) = \phi(x)$ $\forall x \in X$.*

Ejemplo 10. Sea $\alpha = \frac{p}{q}$ con $p, q \in \mathbb{Z}$ tales que $p < q$. Consideremos la transformación traslación del intervalo $T_\alpha : \Pi \to \Pi$ definida como $T_\alpha(x) = x + \alpha$

mód 1. Sea $x \in \Pi$ y consideramos su órbita $A = \{T_\alpha^n(x)\}_{n\in\mathbb{N}}$, es decir,

$$\begin{aligned}
&x \\
&x + \frac{p}{q} \quad \text{mód } 1 \\
&x + 2\frac{p}{q} \quad \text{mód } 1 \\
&\vdots \\
&x + (q-1)\frac{p}{q} \quad \text{mód } 1.
\end{aligned}$$

Como T_α es invertible, entonces verificar que el conjunto A es T_α-invariante es equivalente a verificar que se satisface $T_\alpha(A) = A$, lo cual es claro al observar de manera explícita los elementos de la órbita de x. Más aún, de manera general, se debe observar que cualquier órbita periódica de una transformación invertible T es un conjunto T-invariante.

La siguiente proposición muestra la relación que existe entre los conjuntos T-invariantes y las funciones T-invariantes. La condición necesaria y suficiente para que una función ϕ sea T-invariante es que los conjuntos $\phi^{-1}(\alpha)$ sean conjuntos T-invariantes.

Proposición 13. *Sea (X, S, μ) un espacio de medida finita, $T : X \to X$ una transformación medible y $\phi \in L^1(X, \mu)$. ϕ es una función T-invariante si y sólo si para cada $\alpha \in \mathbb{R}$, se tiene que los conjuntos $\phi^{-1}(\alpha)$ son conjuntos T-invariantes.*

Demostración. Sea $\alpha \in \mathbb{R}$. Por definición, el conjunto $\phi^{-1}(\alpha)$ es un conjunto T-invariante si y sólo si $\phi^{-1}(\alpha) = T^{-1}(\phi^{-1}(\alpha))$, es decir,

$$\begin{gathered}
x \in \phi^{-1}(\alpha) \Leftrightarrow x \in T^{-1}(\phi^{-1}(\alpha)) \\
\phi(x) = \alpha \Leftrightarrow \phi(T(x)) = \alpha.
\end{gathered}$$

Así pues, se tiene $\phi(x) = \phi(T(x))$ para cada $x \in X$. Por lo tanto, $\phi^{-1}(\alpha)$ es un conjunto T-invariante para cada $\alpha \in \mathbb{R}$ si y sólo si ϕ es una función T-invariante. $\square$

En ocasiones es conveniente "relajar" la definición de conjuntos T-invariantes y funciones T-invariantes relativilizándolas con respecto a conjuntos de medida cero.

Definición 19. *Sea (X, S, μ) un espacio de medida finita, $T : X \to X$ una transformación medible. Definimos lo siguiente:*

- *Un conjunto medible $A \subseteq X$ es un* ***conjunto T-invariante c.d. (rel μ)*** *si y sólo si existe un conjunto medible y T-invariante $B \subseteq A$ tal que $\mu(A \setminus B) = 0$.*

- *Una función medible $\phi : X \to \mathbb{R}$ es una* ***función T-invariante c.d. (rel μ)*** *si y sólo si existe un conjunto medible y T-invariante $B \subseteq A$ tal que $\mu(A \setminus B) = 0.$, de manera que la restricción $\phi|_B$ es una función $T|_B$-invariante.*

En la siguiente sección trabajaremos con un tipo particular de función invariante que recibe el nombre de *esperanza condicional* y con los conjuntos T-invariantes que induce en virtud de la Proposición 13.

2.2.3. Esperanza Condicional

La *esperanza condicional* es una noción que tiene su origen en la Probabilidad. La *esperanza* de una *variable aleatoria* ϕ, o bien, simplemente de una función medible ϕ es un punto de equilibrio que se *espera* que adquiera la función ϕ. Se define de manera formal como

$$E(\phi) = \int \phi \mathrm{d}P,$$

donde P es la medida de probabilidad.

La esperanza condicional generaliza la noción de esperanza al calcular el valor esperado únicamente sobre un subconjunto de eventos relevantes. Los eventos en Probabilidad están representados por los conjuntos de la σ-álgebra S. Los eventos relevantes para calcular la esperanza condicional están representados por la **σ-subálgebra** $F \subseteq S$ que es un subconjunto de S con estructura de σ-álgebra. Se define de manera formal como

$$E(\phi|F) = \sum_{G \in F} \int \phi \mathrm{d}P(\phi^{-1}((-\infty, x])|G),$$

donde $P(\phi^{-1}((-\infty, x])|G)$ es la *probabilidad condicional* de $\phi^{-1}((-\infty, x])$ con respecto a $G \in F$.

La principal característica de la esperanza condicional es que cumple lo que se conoce como la *propiedad fundamental de la esperanza condicional*, que es la que la integral de la esperanza de una función integrable ϕ respecto a una σ-subálgebra coincide con la integral de la función ϕ,

$$\int_G E(\phi|F) \mathrm{d}P = \int_G \phi \mathrm{d}P \qquad \forall G \in F.$$

A continuación generalizamos la noción de esperanza condicional dentro un espacio de medida $(X, S\mu)$ para funciones integrables $\phi \in L^1(X, \mu)$. Definimos la esperanza condicional de una función ϕ como una función ϕ_F que satisface la propiedad fundamental de la esperanza condicional para una σ-subálgebra F.

Definición 20. *Sea (X, S, μ) un espacio de medida, $F \subseteq S$ una σ-subálgebra y $\phi \in L^1(X, \mu)$. Definimos la* ***esperanza condicional*** *de ϕ con respecto a F como una función $\phi_F \in L^1(X, \mu)$ que satisface*

$$\int_G \phi_F \mathrm{d}\mu = \int_G \phi \mathrm{d}\mu \qquad \forall G \in F.$$

La siguiente proposición asegura la existencia de la esperanza condicional, es decir, dada una función $\phi \in L^1(X,\mu)$, probaremos que existe la función $\phi_F \in L^1(X,\mu)$ que satisface la propiedad fundamental. Sin embargo, se probará antes un lema auxiliar a la prueba. El lema que se probará echa mano del *Lema Básico de Aproximación* y del *Teorema de la Convergencia Monótona* que enunciamos previamente.

Lema 4. *Sea (X,S,μ) un espacio de medida y $f : X \to \mathbb{R}$ una función medible. Entonces la función $\nu_f : S \to \mathbb{R}$ definida como*

$$\nu_f(E) = \int_E f \mathrm{d}\mu \qquad \forall E \in S,$$

es una medida. Más aun, ν_f es ***absolutamente continua*** *con respecto a μ ($\nu_f \ll \mu$), es decir, para cada $E \in S$, si $\mu(E) = 0$, entonces $\nu_f(E) = 0$.*

Demostración. Para que la función ν_f sea una medida, se debe verificar que satisfaga lo siguiente:

- $\nu_f(\emptyset) = 0$.
- Para cada $E \in S$ se cumple que $\nu_f(E) \geq 0$.
- Si $(E_n)_{n\in\mathbb{N}}$ es una sucesión de conjuntos en S tal que $E_i \cap E_j = \emptyset$ para cada $i, j \in \mathbb{N}$ con $i \neq j$, entonces se tiene que,

$$\nu_f(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \nu_f(E_n).$$

Veamos que se satisface la primera propiedad.

$$\nu_f(\emptyset) = \int_{\emptyset} f \mathrm{d}\mu = \int f\chi_{\emptyset} \mathrm{d}\mu = \int 0 \mathrm{d}\mu = 0.$$

Para ver que se satisface la segunda propiedad, observamos que como $\emptyset \subseteq E$ para cada $E \in S$, entonces se tiene que $\chi_{\emptyset} \leq \chi_E$ y esto implica que $f\chi_{\emptyset} \leq f\chi_E$. Así pues,

$$\int f\chi_{\emptyset} \mathrm{d}\mu \leq \int f\chi_E \mathrm{d}\mu$$
$$\int_{\emptyset} f \mathrm{d}\mu \leq \int_E f \mathrm{d}\mu$$
$$0 \leq \int_E f \mathrm{d}\mu = \nu_f(E).$$

Por lo tanto, para cada $E \in S$ se tiene que $\nu_f(E) \geq 0$. Finalmente, para ver que se satisface la tercera propiedad, consideremos $(E_n)_{n\in\mathbb{N}}$ una sucesión de conjuntos en S tal que $E_i \cap Ej = \emptyset$ para cada $i, j \in \mathbb{N}$. Así pues,

$$\chi_{\cup_{i=1}^{m} E_i} = \sum_{i=1}^{m} \chi_{E_i}$$

para cada número natural m. Definimos la función $g_m : X \to \mathbb{R}$ como,

$$g_m = f\chi_{\cup_{i=1}^{m}(E_i)} \qquad \forall m \in \mathbb{N}.$$

Veamos que la función g_m satisface las siguientes propiedades:

- Para cada número natural m se cumple $g_m \leq g_{m+1}$.
- $g_m \to f\chi_{\cup_{i=1}^{\infty} E_i}$ cuando $m \to \infty$.

La primera propiedad se sigue directamente del hecho de que $\cup_{i=1}^{m} E_i \subseteq \cup_{i=1}^{m+1} E_i$. Esto implica que, $\chi_{\cup_{i=1}^{m} E_i} \leq \chi_{\cup_{i=1}^{m+1} E_i}$ Por lo tanto,

$$g_m \leq g_{m+1} \qquad \forall m \in \mathbb{N}.$$

Veamos que se cumple la segunda propiedad. Sea $x \in X$, entonces se tienen los siguientes casos:

Caso 1: $x \notin \cup_{i=1}^{\infty} E_i$. Entonces $f\chi_{\cup_{i=1}^{\infty} E_i} = 0$.
Así también $x \notin E_i$ para cada $i \in \mathbb{N}$, luego $g_m(x) = f\chi_{\cup_{n=1}^{m} E_n} = 0$ para cada m natural. Por lo tanto $g_m \to f\chi_{\cup_{i=1}^{\infty} E_i}$ si $m \to \infty$.

Caso 2: $x \in \cup_{i=1}^{\infty} E_i$. Entonces $f\chi_{\cup_{i=1}^{\infty} E_i}(x) = f(x)$.
Así también $x \in \cup_{i=1}^{m} E_i$ $\forall m \geq n$. Luego, $g_m(x) = f(x)$ $\forall m \geq n$. Por lo tanto $g_m \to f\chi_{\cup_{i=1}^{\infty} E_i}$ si $m \to \infty$.

Ahora bien, como $g_m \to f\chi_{\cup_{i=1}^{\infty} E_i}$ si $m \to \infty$ y $g_m \leq g_{m+1}$ para cada $m \in \mathbb{N}$, por el Teorema de la Convergencia Monótona tenemos que,

$$\begin{aligned}\nu_f(\cup_{i=1}^{\infty} E_i) &= \int f\chi_{\cup_{i=1}^{\infty} E_i} \mathrm{d}\mu \\ &= \lim_{m\to\infty} \int g_m \mathrm{d}\mu.\end{aligned}$$

Por otra parte, como $\chi_{\cup_{i=1}^{m} E_i} = \sum_{i=1}^{m} \chi_{E_i}$ tenemos que,

$$\begin{aligned}\lim_{m\to\infty} \int f\chi_{\cup_{i=1}^{m} E_i} \mathrm{d}\mu &= \lim_{m\to\infty} \int \sum_{i=1}^{m} f\chi_{E_i} \mathrm{d}\mu \\ &= \lim_{m\to\infty} \sum_{i=1}^{m} \int f\chi_{E_i} \mathrm{d}\mu \\ &= \lim_{m\to\infty} \sum_{i=1}^{m} \nu_f(E_i) \\ &= \sum_{i=1}^{\infty} \nu_f(E_i).\end{aligned}$$

Por lo tanto, ν_f es una medida. Falta probar que que $\nu_f \ll \mu$. Sea $E \in S$ tal que $\mu(E) = 0$. Veamos que $\nu_f(E) = 0$.

$$\nu_f(E) = \int_E f \mathrm{d}\mu = \int f\chi_E \mathrm{d}\mu.$$

Por hipótesis se tiene que f es una función medible. Por el Lema Básico de Aproximación sabemos que existe una sucesión $(s_n)_{n\in\mathbb{N}}$ creciente de funciones simples tales que

$$\lim_{n\to\infty} s_n = f.$$

Las funciones simples se pueden ver de la siguiente manera,

$$s_n = \sum_{i=1}^{k_n} \alpha_{i,n}\chi_{E_{i,n}}.$$

Por el Teorema de la Convergencia Monótona se sigue que,

$$\begin{aligned}
\int f\chi_E \mathrm{d}\mu &= \int \lim_{n\to\infty} s_n\chi_E \mathrm{d}\mu \\
&= \lim_{n\to\infty} \int \sum_{i=1}^{k_n} \alpha_{i,n}\chi_{E_{i,n}}\chi_E \mathrm{d}\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^{k_n} \alpha_{i,n} \int \chi_{E_{i,n}}\chi_E \mathrm{d}\mu \\
&= \lim_{n\to\infty} \sum_{i=1}^{k_n} \alpha_{i,n} \int \chi_{E_{i,n}\cap E} \mathrm{d}\mu.
\end{aligned}$$

Observamos que como $\chi_{E_{i,n}\cap E} \leq \chi_E$,

$$0 \leq \int \chi_{E_{i,n}\cap\chi_E} \mathrm{d}\mu \leq \int \chi_E \mathrm{d}\mu = \mu(E) = 0.$$

Esto implica que $\int \chi_{E_{i,n}\cap\chi_E}\mathrm{d}\mu = 0$ para cada $n \in \mathbb{N}$ y por lo tanto se concluye que $\nu_f(E) = 0$, es decir, que $\nu_f \ll \mu$. □

Para probar la existencia de la esperanza condicional, ocuparemos el *Teorema de Radón-Nikodym*, mismo que enunciamos a continuación.

Teorema 16 (Teorema de Radón-Nikodym)**.** *Sea (X, S) un espacio medible y sean μ, ν medidas σ-finitas tales que $\nu \ll \mu$. Entonces existe una única $f \in M(X, S)$ tal que,*

$$\nu(E) = \int_E f\mathrm{d}\mu \qquad \forall E \in S.$$

Proposición 14. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$ y sea $F \subset S$ una σ-subálgebra. Para cada función $\phi \in L^1(X, \mu)$ con respecto a la σ-álgebra S, existe una función $\phi_F \in L^1(X, \mu)$ con respecto a la σ-subálgebra F tal que,*

$$\int_G \phi_F d\mu = \int_G \phi d\mu \qquad \forall G \in F.$$

Demostración. Sea $\phi \in L^1(X, \mu)$ con respecto a la σ-álgebra S. Definimos la función $\nu_\phi : X \to \mathbb{R}$ como

$$\nu_\phi(G) = \int_G \phi \mathrm{d}\mu \qquad \forall G \in F.$$

Por el Lema 4 sabemos que ν_ϕ es una medida tal que $\nu_\phi \ll \mu$. Así pues, por el Teorema de Radón-Nikodym, se sigue que existe una función medible $\phi_F \in L^1(X, \mu)$ con respecto a la σ-subálgebra F tal que

$$\nu_\phi(G) = \int_G \phi_F \mathrm{d}\mu \qquad \forall G \in F.$$

Por lo tanto,

$$\int_G \phi \mathrm{d}\mu = \int_G \phi_F \mathrm{d}\mu \qquad \forall G \in F.$$

□

La proposición anterior garantiza la existencia de la esperanza condicional para una función $\phi \in L^1(X, \mu)$ dada. Sin embargo, no se garantiza que sea única, pues puede cambiar sobre cualquier conjunto de medida cero en la σ-subálgebra F. La siguiente proposición muestra una propiedad auxiliar que tiene la Esperanza Condicional de una función integrable.

Proposición 15. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$, sea $F \subset S$ una σ-subálgebra y sea $\phi \in L^1(X, \mu)$. Entonces la esperanza condicional ϕ_F de ϕ cumple*

$$-\phi_F = (-\phi)_F \qquad \textit{c.d. (rel } \mu).$$

Demostración. Por hipótesis, $\phi \in L^1(X, \mu)$. Esto implica que $-\phi \in L^1(X, \mu)$. Por el Teorema 14, existe la esperanza condicional de $-\phi$ con respecto a F y además

$$\int_G (-\phi)_F \mathrm{d}\mu = \int_G -\phi \mathrm{d}\mu \qquad \forall G \in F.$$

Por otra parte, se tiene que

$$\int_G -\phi \mathrm{d}\mu = -\int_G \phi \mathrm{d}\mu = -\int_G \phi_F \mathrm{d}\mu = \int_G -\phi_F \mathrm{d}\mu \qquad \forall G \in F.$$

Así pues,

$$\int_G (-\phi)_F \mathrm{d}\mu \int_G -\phi_F \mathrm{d}\mu \qquad \forall G \in F.$$

Por lo tanto, $(-\phi)_F = -\phi_F$ c.d. (rel μ). □

El siguiente resultado muestra que, dado un espacio de medida finito y una transformación medible T, el conjunto de los conjuntos T-invariantes constituye una σ-subálgebra. La σ-subálgebra de los conjuntos T-invariantes es importante porque el Teorema Ergódico de Birkhoff afirma que la sucesión de promedios converge precisamente a la esperanza condicional de una función ϕ integrable con respecto a la σ-subálgebra de los conjuntos T-invariantes. Más aun, probaremos que la esperanza condicional de una función ϕ con respecto a la σ-subálgebra de los conjuntos T-invariantes es una función T-invariante.

Proposición 16. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$ y $T : X \to X$ una transformación medible. Entonces el conjunto de los conjuntos T-invariantes,*

$$F = \{E \in S : T^{-1}(E) = E\},$$

es una σ-subálgebra.

Demostración. Por definición de F, se sigue que $F \subseteq S$. Así pues, sólo hay que probar que F es una σ-álgebra, es decir, que satisface las siguientes propiedades:

- $E, G \in F$, implica que $E \setminus G \in F$.
- Si $(G_n)_{n\in\mathbb{N}}$ es una sucesión de elementos de F, entonces $\bigcup_{n=1}^{\infty} G_n \in F$.
- $X \in F$.

Sea $E, G \in F$, esto implica que $T^{-1}(E) = E$ y $T^{-1}(G) = G$. Por propiedades de la imagen inversa se tiene,

$$E \setminus G = T^{-1}(E) \setminus T^{-1}(G) = T^{-1}(E \setminus G).$$

Por lo tanto, $E \setminus G \in F$. Veamos que se satisface la segunda propiedad. Sea $(G_n)_{n\in\mathbb{N}}$ una sucesión de elementos de F. Esto implica que $G_n = T^{-1}(G_n)$ para cada $n \in \mathbb{N}$. Asimismo, por propiedades de la imagen inversa se tiene que,

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} T^{-1}(E_n) = T^{-1}\left(\bigcup_{n=1}^{\infty} E_n\right).$$

Por lo tanto $\bigcup_{n=1}^{\infty} E_n \in F$. Finalmente, para probar la última propiedad, observamos que por hipótesis S es una σ-álgebra. Entonces tenemos que $X \in S$ y también por hipótesis tenemos que la transformación $T : X \to X$ es una transformación medible. Esto implica que $T^{-1}(X) = X$. Por lo tanto, podemos concluir que F es una σ-subálgebra. □

La siguiente proposición muestra una equivalencia de que una función $\phi \in L^1(X, \mu)$ sea una función T-invariante. Una función integrable ϕ es una función T-invariante si y sólo si es una función medible con respecto a la σ-subálgebra de los conjuntos T-invariantes.

Proposición 17. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$, sea $T : X \to X$ una transformación medible y $\phi \in L^1(X, \mu)$. Consideremos la σ-subálgebra $F \subset S$ de los conjuntos T-invariantes. Entonces, ϕ es una función F-medible si y sólo si ϕ es una función T-invariante.*

Demostración. Por hipótesis se tiene que $\phi \in L^1(X, \mu)$. Para probar el enunciado de la proposición, se probará que la función ϕ es F-medible si y sólo si $\phi^{-1}(\alpha)$ es un conjunto T-invariante para cada $\alpha \in \mathbb{R}$ y posteriormente se ocupará la Proposición 13.

Supongamos primero que ϕ es F-medible. Como el conjunto unitario $\{\alpha\} \subseteq \mathbb{R}$ es medible para cada $\alpha \in \mathbb{R}$, entonces tiene que $\phi^{-1}(\alpha) \in F$ y por definición de F se tiene que,

$$(\phi \circ T)^{-1}(\alpha) = \phi^{-1}(\alpha) \qquad \forall \alpha \in \mathbb{R}.$$

Por lo tanto, $\phi^{-1}(\alpha)$ es un conjunto T-invariante para cada $\alpha \in \mathbb{R}$.

Recíprocamente, supongamos que $\phi^{-1}(\alpha)$ es un conjunto T-invariante para cada $\alpha \in \mathbb{R}$ y veamos que ϕ es una función F-medible. Sea $B \subseteq \mathbb{R}$ medible. Veamos que $\phi^{-1}(B) \in F$, es decir, por definición de F, que $\phi^{-1}(B) = (\phi \circ T^{-1})(B) \in F$.

$$\begin{aligned}
(\phi \circ T)^{-1}(B) &= (\phi \circ T)^{-1}\left(\bigcup_{\alpha \in B} \{\alpha\}\right) \\
&= \bigcup_{\alpha \in B} (\phi \circ T)^{-1}(\alpha) \\
&= \bigcup_{\alpha \in B} \phi^{-1}(\alpha) \\
&= \phi^{-1}\left(\bigcup_{\alpha \in B} (\alpha)\right) \\
&= \phi^{-1}(B).
\end{aligned}$$

Por lo tanto, ϕ es F-medible. Finalmente, por la Proposición 13, concluimos que ϕ es medible si y sólo si es T-invariante. □

Una consecuencia directa de la proposición anterior es que en particular la esperanza condicional de una función ϕ con respecto a la σ-subálgebra de los conjuntos T-invariantes es una función T-invariante.

Corolario 1. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$, sea $T : X \to X$ una transformación medible y $\phi \in L^1(X, \mu)$. Sea también $F \subset S$ la σ-subálgebra de los conjuntos T-invariantes. Entonces, la esperanza condicional ϕ_F de la función ϕ con respecto a F es una función T-invariante.*

Demostración. Se sigue directo de la Proposición 17. □

Para terminar la sección, se calculará de manera explícita la esperanza condicional de una función integrable con respecto a una σ-subálgebra.

Ejemplo 11. Sea (X, S, μ) un espacio de medida y sea $B \in S$. Consideramos $\phi = \chi_B$ que se sabe que es una función integrable. Tomamos un conjunto $E \in S$ tal que $\mu(E) > 0$ y también $\mu(X \setminus E) > 0$. Consideramos la siguiente σ-subálgebra,

$$F = \{\emptyset, E, X \setminus E, X\}.$$

Por la Proposición 14 sabemos que existe la esperanza condicional de χ_B y por definición, sabemos que satisface,

$$\int_B \phi_F \mathrm{d}\mu = \int_B \chi_E \mathrm{d}\mu = \mu(E \cap B) \qquad \forall B \in F.$$

Observamos que para que ϕ_F sea una función medible, se necesita que sea constante sobre E y sobre $X \setminus E$. Como ϕ_F satisface la propiedad fundamental se puede inferir que se debe definir como,

$$\phi_F(x) = \begin{cases} \frac{\mu(E\cap B)}{\mu(E)} & si\ x \in E \\ \frac{\mu((X\setminus E)\cap B)}{\mu(X\setminus E)} & si\ x \in X \setminus E. \end{cases}$$

El ejemplo anterior es importante por las aplicaciones que se utilizan en Probabilidad. La σ-subálgebra que se define permite distinguir el hecho de que ocurra o no un evento y si se toma la medida μ como la probabilidad. Observamos que la esperanza condicional se puede escribir en términos de probabilidad condicional como sigue,

$$E(X|F) = P(B|E)\chi_E + P(B|X \setminus E)\chi_{X\setminus E}.$$

En [19] se afirma que este hecho da una caracterización ulterior de la esperanza condicional como una generalización de la probabilidad condicional.

2.2.4. Teorema Ergódico de Birkhoff

En esta sección se presenta la prueba del Teorema Ergódico de Birkhoff. Probaremos que la frecuencia con la que permanece la órbita de un punto evaluado en una función integrable ϕ en un conjunto $E \subseteq X$ medible, coincide con su valor evaluado en la esperanza condicional de ϕ con respecto a la σ-subálgebra de los conjuntos T-invariantes.

En la prueba del Teorema Ergódico, utilizaremos el Teorema de la Convergencia Dominada, mismo que enunciamos sin prueba a continuación.

Teorema 17 (Convergencia Dominada)**.** *Sea (X, S, μ) un espacio de medida y (f_n) una sucesión de funciones medible tales que,*

$$f_n(x) \to f(x) \qquad c.d.\ rel(\mu).$$

Para alguna función medible f. Supongamos que existe $g \in L^1(X, \mu)$ tal que $|f_n| \leq g$ c.d. rel(μ) para cada natural n. Entonces,

- $f \in L^1(X, \mu)$.
- $$\int_E f \mathrm{d}\mu = \lim_{n\to\infty} \int_E f_n \mathrm{d}\mu \qquad \forall E \in S.$$

Teorema 18 (Ergódico de Birkhoff)**.** *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$ y sea $T : X \to X$ una transformación que preserva la medida μ. Si $\phi \in L^1(X, \mu)$, entonces el límite*

$$\phi_T(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x))$$

existe para casi toda $x \in X$. Más aún, la función ϕ_T que define satisface las siguientes propiedades:

- *ϕ_T es una función T-invariante c.d. (rel μ).*
- *$\phi_T \in L^1(X, \mu)$.*

Además,

$$\int \phi_T \mathrm{d}\mu = \int \phi \mathrm{d}\mu.$$

Demostración. Consideremos el conjunto de conjuntos T-invariantes de X,

$$F = \{A \in S : T^{-1}(A) = A\}. \tag{2.1}$$

Por la Proposición 16, sabemos que F es una σ-subálgebra. Por la Proposición 14 y como por hipótesis $\phi \in L^1(X, \mu)$, se tiene que existe la esperanza condicional de ϕ, es decir, una función $\phi_F \in L^1(X, \mu)$ tal que

$$\int_A \phi_F d\mu = \int_A \phi d\mu \qquad A \in F.$$

Sea $\epsilon > 0$, definimos $\psi = \phi - \phi_F - \epsilon$. Integramos,

$$\begin{aligned} \int_A \psi d\mu &= \int_A \phi d\mu - \int_A \phi_F d\mu - \epsilon \int_A d\mu & \qquad \forall A \in F \\ &= -\epsilon\mu(A) & \forall A \in F. \end{aligned}$$

Observamos que si $\int_A \psi d\mu \geq 0$, entonces $-\epsilon\mu(A) \geq 0$. Sin embargo, $0 \leq -\epsilon\mu(A) \leq 0$. Por lo tanto,

$$\begin{aligned} -\epsilon\mu(A) &= 0 \\ \mu(A) &= 0. \end{aligned}$$

Es decir, para probar que los conjuntos T-invariantes de X tienen medida cero, basta probar que $\int_A \psi d\mu \geq 0$. Dada $\psi \in L^1(X, \mu)$, definimos la siguiente familia de funciones. Para cada $n \in \mathbb{N}$ definimos

$$\psi_n = \max \left\{ \sum_{k=0}^{l-1} \psi \circ T^k : 1 \leq l \leq n \right\}.$$

Veamos que $\psi_{n+1} \geq \psi_n$ $\forall n \in \mathbb{N}$. Por inducción sobre n. Para $n = 1$,

$$\begin{aligned}
\psi_1 &= \text{máx}\left\{\sum_{k=0}^{l-1} \psi \circ T^k : 1 \leq l \leq 1\right\} \\
&= \text{máx}\left\{\sum_{k=0}^{l-1} \psi \circ T^k : l = 1\right\} \\
&= \text{máx}\left\{\sum_{k=0}^{0} \psi \circ T^k\right\} \\
&= \text{máx}\{\psi\} \\
&= \psi.
\end{aligned}$$

Por otra parte,

$$\begin{aligned}
\psi_2 &= \text{máx}\left\{\sum_{k=0}^{l-1} \psi \circ T^k : 1 \leq l \leq 2\right\} \\
&= \text{máx}\{\psi, \psi + \psi \circ T\} \geq \psi = \psi_1.
\end{aligned}$$

Por lo tanto $\psi_2 \geq \psi_1$. Por hipótesis de inducción, supongamos que $\psi_n \geq \psi_k$ $\forall k \in \{1, 2, \cdots, n-1\}$. Para verificar el paso inductivo, veamos que $\psi_{n+1} \geq \psi_n$.

$$\begin{aligned}
\psi_{n-1} &= \text{máx}\left\{\sum_{k=0}^{l-1} \psi \circ T^k : 1 \leq l \leq n+1\right\} \\
&= \{\psi, \psi + \psi \circ T, \cdots, \psi, \psi + \psi \circ T + \cdots + \psi \circ T^{n+1}\} \\
&\geq \{\psi, \psi + \psi \circ T, \cdots, \psi, \psi + \psi \circ T + \cdots + \psi \circ T^{n}\} \\
&= \psi_n.
\end{aligned}$$

Por lo tanto, $\psi_{n+1} \geq \psi_n$ para cada $n \in \mathbb{N}$. Ahora bien, veamos que $(\psi_n)_{n\in\mathbb{N}} \subseteq L^1(X, \mu)$. Por la desigualdad del triángulo, observamos que $|\psi_n| \leq \sum_{k=0}^{l-1} |\psi \circ T^k|$ y por la linealidad de la integral, tenemos lo siguiente,

$$\int |\psi_n| d\mu \leq \int \sum_{k=0}^{l-1} |\psi \circ T^k| d\mu = \sum_{k=0}^{l-1} \int |\psi \circ T^k| d\mu.$$

Como $\psi \in L^1(X, \mu)$, entonces también $|\psi| \in L^1(X, \mu)$. Como μ es una medida finita y T preserva la medida μ, por el Teorema 15, tenemos que $|\psi \circ T^k| \in L^1(X, \mu)$ para cada $k \in \mathbb{N}$. Ademas,

$$\int |\psi \circ T^k| d\mu = \int |\psi| d\mu \qquad \forall k \in \mathbb{N}.$$

Entonces,

$$\sum_{k=0}^{n-1} \int |\psi \circ T^k| d\mu = \sum_{k=0}^{n-1} \int |\psi| d\mu = n \int |\psi| d\mu < +\infty.$$

Por lo tanto $\int |\psi_n| d\mu < +\infty \quad \forall n \in \mathbb{N}$. Esto implica que $(\psi_n)_{n\in\mathbb{N}} \subseteq L^1(X,\mu)$. Más aún, observamos lo siguiente,

$$\begin{aligned}\psi_{n+1} &= \text{máx}\left\{\sum_{k=0}^{l-1} \psi \circ T^k : 1 \leq l \leq n+1\right\} \\ &= \text{máx}\left\{\sum_{k=0}^{l} \psi \circ T^k : 1 \leq l \leq n\right\} \\ &= \psi + \text{máx}\left\{0, \sum_{k=1}^{l} \psi \circ T^k : 1 \leq l \leq n\right\} \\ &= \psi + \text{máx}\{0, \psi_n \circ T\}.\end{aligned}$$

Así pues, $\lim_{n\to\infty} \psi_{n+1}(x) = +\infty$ si y sólo si $\lim_{n\to\infty} \psi_n(T(x)) = +\infty$. Este hecho nos ayuda a probar que el conjunto de puntos en los que $\lim_{n\to\infty} \psi_n(x) = +\infty$ es un conjunto T-invariante y por ende, será de medida cero. Veamos que $A = \{x \in X : \lim_{n\to\infty} \psi_n(x) = +\infty\}$ es un conjunto T-invariante, es decir, que $T^{-1}(A) = A$.

$$\begin{aligned}x \in T^{-1}(A) &\Leftrightarrow T(x) \in A \\ &\Leftrightarrow \lim_{n\to\infty} \psi_n(T(x)) = +\infty \\ &\Leftrightarrow \lim_{n\to\infty} \psi_{n+1}(x) = +\infty \\ &\Leftrightarrow \lim_{n\to\infty} \psi_n(x) = +\infty \\ &\Leftrightarrow x \in A.\end{aligned}$$

Por lo tanto $T^{-1}(A) = A$. Como,

$$\psi_{n+1} = \psi + \text{máx}\{0, \psi_n \circ T\}.$$

Entonces,

$$\begin{aligned}\psi_{n+1} - \psi_n \circ T &= \psi + \text{máx}\{0, \psi_n \circ T\} - \psi_n \circ T \\ &= \psi - \text{mín}\{0, \psi_n \circ T\}.\end{aligned}$$

Tenemos lo siguiente,

$$\begin{aligned}\lim_{n\to\infty} (\psi_{n+1} - \psi_n \circ T) &= \lim_{n\to\infty} (\psi - \text{mín}\{0, \psi_n \circ T) \\ &= \psi - \lim_{n\to\infty} (\text{mín}\{0, \psi_n \circ T\}) \\ &= \psi - 0 \\ &= \psi.\end{aligned}$$

Por lo tanto, $\psi_{n+1} - \psi_n \circ T \searrow \psi$ cuando $n \to \infty$ en A. Ahora veamos que $0 \leq \psi_{n+1} - \psi_n \circ T - \psi \leq \psi_2 - \psi_1 \circ T - \psi$. Por inducción sobre n. Tomamos como base de la inducción a $n = 1$.

$$0 \leq \psi_2 - \psi_1 \circ T - \psi \leq \psi_2 - \psi_1 \circ T - \psi.$$

Por hipótesis de inducción, supongamos que se cumple para $n \in \mathbb{N}$, es decir, que se satisface,

$$0 \leq \psi_{n+1} - \psi_n \circ T - \psi \leq \psi_2 - \psi_1 \circ T - \psi.$$

Para verificar el paso inductivo, veamos que $0 \leq \psi_{n+2} - \psi_{n+1} \circ T - \psi \leq \psi_2 - \psi_1 \circ T - \psi$. Sabemos,

$$\psi_{n+2} - \psi_{n+1} \circ T = \psi - \text{mín}\{0, \psi_{n+1} \circ T\}.$$

Entonces,

$$\begin{aligned} \psi_{n+2} - \psi_{n+1} \circ T - \psi &= \text{mín}\{0, \psi_{n+1} \circ T\} \\ &= \text{máx}\{0, \psi_{n+1} \circ T\} - \psi_{n+1} \circ T \\ &= \psi_{n+1} - \psi_n \circ T - \psi. \end{aligned}$$

Por la hipótesis de inducción, tenemos que $\psi_{n+1} - \psi_n \circ T - \psi \leq \psi_2 - \psi_1 \circ T - \psi$. Por lo tanto, $\psi_{n+1} - \psi_n \circ T - \psi \leq \psi_2 - \psi_1 \circ T - \psi \ \forall n \in \mathbb{N}$. Como $\psi_2, \psi_1 \circ T, \psi \in L^1(X, \mu)$, entonces también $\psi_2 - \psi_1 \circ T - \psi, \psi_2 - \psi_1 \circ T \in L^1(X, \mu)$. Asimismo, como $\psi_{n+1}, \psi_n \circ T, \psi \in L^1(X, \mu) \ \forall n \in \mathbb{N}$, entonces,

$$\begin{aligned} \psi_{n+1} - \psi_n \circ T - \psi \in L^1(X, \mu) \qquad & \forall n \in \mathbb{N} \\ \psi_{n+1} - \psi_n \circ T \in L^1(X, \mu) \qquad & \forall n \in \mathbb{N}. \end{aligned}$$

Como μ es finito y T-invariante en X, por el Teorema 15 se tiene que $\psi_n \circ T \in L^1(X, \mu)$ y también

$$\int \psi_n \circ T d\mu = \int \psi_n d\mu \qquad \forall n \in \mathbb{N}.$$

Por la linealidad de la integral y por el Teorema 15 tenemos

$$\begin{aligned} 0 \leq \int_A (\psi_{n+1} - \psi_n) d\mu &= \int_A \psi_{n+1} d\mu - \int_A \psi_n d\mu \qquad \forall n \in \mathbb{N} \\ &= \int_A \psi_{n+1} d\mu - \int_A \psi_n \circ T d\mu \qquad \forall n \in \mathbb{N} \\ &= \int_A (\psi_{n+1} - \psi_n \circ T) d\mu \qquad \forall n \in \mathbb{N}. \end{aligned}$$

Además como $|\psi_{n+1} - \psi_n \circ T| \leq \psi_2 - \psi_1 \circ T$ y también $\psi_{n+1} - \psi_n \circ T \searrow \psi$ cuando $n \to +\infty$. Por el Teorema de la Convergencia Dominada,

$$\begin{aligned} \lim_{n\to\infty} \int_A (\psi_{n+1} - \psi_n \circ T) d\mu &= \int_A \lim_{n\to\infty} (\psi_{n+1} - \psi_n \circ T) d\mu \\ &= \int_A \psi d\mu. \end{aligned}$$

Por lo tanto, $\int_A \psi d\mu \geq 0$. Como A es un conjunto T-invariante y sabemos que los conjuntos T-invariantes tienen medida cero, entonces se tiene que $\lim_{n\to\infty} \psi_n <$

$+\infty$ c.d. (rel μ). Ahora bien, por la manera en que se definió ψ_n, sabemos que $\sum_{k=0}^{n-1} \psi \circ T^k \leq \psi_n$. Entonces se tiene,

$$\frac{1}{n}\sum_{k=0}^{n-1} \psi \circ T^k \leq \frac{1}{n}\psi_n.$$

También,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \psi \circ T^k \leq \lim_{n\to\infty} \frac{1}{n}\psi_n = 0 \qquad \text{c.d. (rel}\mu).$$

Por la Proposición 1, sabemos que ϕ_F es T-invariante. Entonces,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \psi \circ T^k = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (\phi \circ T^k - \phi_F \circ T^k - \epsilon).$$

Como ϕ_F es T-invariante, entonces $\phi_F \circ T^k = \phi_F$. Asimismo, también tenemos que,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \psi \circ T^k \leq \phi_F + \epsilon \qquad \text{c.d. (rel } \mu).$$

Observamos que por la Proposición 15 $(-\phi_F) = (-\phi)_F$ y reemplazando ϕ por $-\phi$ se tiene lo siguiente,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k \geq \phi_F - \epsilon \qquad \text{c.d. (rel } \mu).$$

Tenemos que para cada $\epsilon > 0$, $\exists X_\epsilon \subseteq X$ tal que $\mu(X_\epsilon) = \mu(X)$ sobre el cual se tiene

$$\phi_F - \epsilon \leq \liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k \leq \limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k \leq \phi_F + \epsilon.$$

Así pues, sobre el conjunto

$$\bigcap_{m=1}^{\infty} X_{\frac{1}{m}} \subseteq X.$$

Concluimos que $\lim\limits_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \psi \circ T^k = \phi_F$. Finalmente,

$$\phi_T = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k.$$

Entonces $\phi_T = \phi_F$ c.d. (rel μ). Como $\phi_F \in L^1(X,\mu)$, entonces $\phi_T \in L^1(X,\mu)$ y por la Proposición 14, como $\int_A \phi_F d\mu = \int_A \phi d\mu$ para cada $A \in F$, podemos concluir

$$\int \phi_T d\mu = \int \phi_F d\mu = \int \phi d\mu.$$

□

Capítulo 3

Ergodicidad

En este capítulo se introduce la noción de *ergodicidad*. La noción de ergodicidad describe un comportamiento importante de las órbitas de los puntos en un sistema dinámico. De manera general, la ergodicidad captura la idea de que todas las órbitas en un sistema dinámico (salvo un conjunto de medida cero) tocan en algún momento cada conjunto medible. Se verá que la ergodicidad nos permite describir un comportamiento general de los puntos en el Teorema Ergódico de Birkhoff, pues se probará que las sumas de Cèsaro convergen siempre a una constante. Posteriormente se verán algunos ejemplos importantes. Finalmente se desarrollará la noción de *ergodicidad única*. La ergodicidad única es cuando hay una única medida que es preservada por una transformación T. La ergodicidad única tiene a su vez consecuencias sobre el Teorema Ergódico de Birkhoff al asegurar que la convergencia es uniforme.

3.1. Transformaciones Ergódicas

A continuación definimos de manera formal la noción de *ergodicidad* y probaremos las equivalencias más importantes. La palabra "ergódico" fue introducida por primera vez por Boltzmann en sus trabajos sobre *Mecánica Estadística* de 1870 y 1884 [22]. La idea original surgió al considerar grandes sistemas de partículas que interactuaban unas con otras. Por ejemplo, las partículas de un gas encerradas en un recipiente, moviéndose e interactuando unas con otras. La hipótesis que se formuló era que todo sistema de partículas llegaría a un punto de equilibrio después de que pasara un periodo largo de tiempo. Por ejemplo, si consideramos un cilindro con un émbolo que contiene partículas de algún gas en un estado de equilibrio y después movíamos el émbolo, después de un tiempo grande, se tendrá que las partículas alcanzarán a su vez un estado de equilibrio. Más aún, se conjeturó la ***hipótesis ergódica*** como sigue: *Para grandes sistemas de partículas interactuando en equilibrio, las medias temporales están próximas de las medias espaciales.*

Las *medias temporales* son los promedios de los valores observados de una

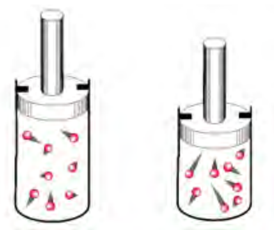


Figura 3.1: Particulas de gas alcanzando el equilibrio.

función $\phi : X \to \mathbb{R}$, al pasar el tiempo. El paso del tiempo se expresa mediante la aplicación sucesiva de una transformación $T : X \to X$. Así pues, la sucesión de los promedios se expresa de la siguiente manera:

$$(s_n)_{n\in\mathbb{N}} = \left(\frac{1}{n}\sum_{k=1}^{n}\phi(T^k(x))\right)_{n\in\mathbb{N}}.$$

Las *medias espaciales* consisten en las integrales respecto a una medida de los valores observados registrados por la función ϕ.

$$\frac{1}{\mu(X)}\int \phi \mathrm{d}\mu.$$

Una formulación formal y precisa de la hipótesis ergódica es

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\phi(T^k(x)) = \frac{1}{\mu(X)}\int \phi \mathrm{d}\mu.$$

No sólo los sistemas de partículas alcanzan un estado de equilibrio al haber transcurrido un tiempo suficientemente grande, sino que las medias temporales para cada punto coinciden con las medias espaciales. Una de las consecuencias más importantes de que se cumpla la hipótesis ergódica es que el tiempo medio no depende del punto inicial. El valor de la media espacial es un valor constante sin importar el punto que se tome en el espacio X. Las transformaciones $T\colon X \to X$ para las que se satisfaga la hipótesis ergódica serán aquellas a las que llamaremos *transformaciones ergódicas*.

Antes de dar la definición formal de una transformación ergódica, vale la pena hacer una caracterización de los espacios que satisfacen la hipótesis ergódica. Consideremos un espacio de medida (X, S, μ) tal que $\mu(X) = 1$ y un conjunto medible $B \subseteq X$ que sea T-invariante y de *medida positiva*, es decir, se satisface $T^{-1}(B) = B$ y $0 < \mu(B) < 1$. Ahora bien, como B es un conjunto medible, T-invariante y de medida positiva, entonces su complemento también es un conjunto medible, T-invariante y de medida positiva. En efecto, basta observar que, por propiedades de la imagen inversa, se tiene que $T^{-1}(X \setminus B) = X \setminus (T^{-1}(B)) = X \setminus B$. El espacio X se puede ver como la unión disjunta de conjuntos medibles, T-invariantes y de medida positiva de la siguiente manera,

$$X = B \cup (X \setminus B).$$

En este caso, decimos que el espacio X es un ***espacio descomponible***. A continuación mostraremos que cuando el espacio X es un espacio descomponible, no se puede satisfacer la hipótesis ergódica. Más aún, se verá que la hipótesis ergódica se satisface si y sólo si el espacio X es *indescomponible*. Esta caracterización es importante porque nos da a entender que la órbita de cada punto recorre todos los subconjuntos medibles del espacio.

Lo primero que observaremos es que cuando el espacio X es descomponible, el valor del tiempo medio depende del punto que se tome en el espacio. Consideremos una transformación medible $T : X \to X$ y un conjunto $B \subseteq X$ medible, T-invariante y de medida positiva. Tomamos la función característica χ_B, la cual sabemos que es integrable. Veremos que la media temporal difiere de la media espacial, pues la media temporal depende del conjunto T-invariante donde se encuentre el punto. Por una parte, si se toma un punto x en el conjunto B, como $\mu(B) < 1$, entonces se tiene

$$\lim_{n\to\infty} \frac{1}{n} \sum_{n=0}^{n-1} \chi_B(T^k(x)) = 1 \neq \frac{1}{\mu(X)} \int \chi_B d\mu = \frac{\mu(B)}{\mu(X)}.$$

Por otra parte, si se toma y en el conjunto $X \setminus B$, como $\mu(X \setminus B) > 0$, entonces se tiene lo siguiente,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_B(T^k(y)) = 0 \neq \frac{1}{\mu(X)} \int \chi_B d\mu = \frac{\mu(B)}{\mu(X)}.$$

Por lo tanto, la media temporal difiere de la media espacial y depende del punto que se tome. Conversamente, consideremos una función $\phi \in L^1(X, \mu)$. Por el Teorema Ergódico de Birkhoff, existe la media temporal $\phi_T \in L^1(X, \mu)$ y es T-invariante. Observamos que a menos que la función ϕ_T sea constante, el espacio X es un espacio descomponible. En efecto, sea $B \subseteq X$ definido como,

$$B = \{x : \phi_T(x) < a\}$$

Para alguna $a \in \mathbb{R}$ tal que $\mu(B) > 0$. Como ϕ_T es T-invariante, por la Proposición 13, se tiene que B es un conjunto medible, T-invariante y de medida positiva. Por lo tanto, el espacio X es un espacio descomponible. Así pues, para que se cumpla la hipótesis ergódica, se requiere que dado un espacio X, los conjuntos medibles y T-invariantes midan 0 o 1. Siguiendo esta idea, definimos a continuación de manera formal una transformación ergódica.

Definición 21. *Sea (X, S, μ) un espacio medible y $T : X \to X$ una transformación medible. Decimos que la transformación T es una **transformación ergódica** si para todo conjunto $B \subseteq X$ medible y T-invariante se tiene que $\mu(B) = 0$, o bien, $\mu(X \setminus B) = 0$.*

Observamos que en el caso en que el espacio X es de medida finita, o bien, que $\mu(X) = 1$, la condición para que la transformación T sea ergódica es equivalente a pedir que los conjuntos medibles y T-invariantes midan 0 o 1. A continuación veremos la primera equivalencia de la ergodicidad en términos de funciones T-invariantes. Una transformación T es ergódica si y sólo si las funciones T-invariantes son constantes. Antes de probar el criterio de ergodicidad, se probará un lema auxiliar en la prueba.

Lema 5. *Sea (X, S, μ) un espacio de medida, $T : X \to X$ una transformación medible y $\phi : X \to \mathbb{R}$ una función medible y T-invariante c.d. (rel μ). Si $A \subseteq \mathbb{R}$ es un conjunto medible, entonces $\phi^{-1}(A)$ es un conjunto T-invariante c.d. (rel μ).*

Demostración. Para ver que el conjunto $\phi^{-1}(A)$ es un conjunto T-invariante c.d. (rel μ) se debe verificar que $T^{-1}(\phi^{-1}(A)) = \phi^{-1}(A)$ c.d. (rel μ). Así pues, por la Proposición 13 y por propiedades de la imagen inversa, se tiene

$$\begin{aligned} T^{-1}(\phi^{-1}(A)) &= T^{-1}\left(\phi^{-1}\left(\bigcup_{\alpha\in A}\{\alpha\}\right)\right) \\ &= \bigcup_{\alpha\in A} T^{-1}(\phi^{-1}(\{\alpha\})) \\ &= \bigcup_{\alpha\in A} \phi^{-1}(\{\alpha\}) \\ &= \phi^{-1}\left(\bigcup_{\alpha\in A}\{\alpha\}\right) \\ &= \phi^{-1}(A). \end{aligned}$$

□

El criterio de ergodicidad que se muestra a continuación afirma que una transformación es ergódica si cada función T-invariante es una función constante.

Teorema 19 (Criterio de ergodicidad)**.** *Sea (X, S, μ) un espacio de medida y sea $T : X \to X$ una transformación medible. La transformación T es ergódica si y sólo si cada función medible $\phi : X \to \mathbb{R}$ que es T-invariante c.d. (rel μ) es una función constante c.d. (rel μ).*

Demostración. Supongamos primero que T es una transformación ergódica y que $\phi : X \to \mathbb{R}$ es una función medible y T-invariante c.d. (rel μ). Para ver que ϕ es una función constante c.d. (rel μ), se definirá la siguiente familia de conjuntos $\{A_{nm}\}$. Para cada $n \in \mathbb{Z}$ y $m \in \mathbb{N}$ definimos el conjunto A_{nm} como

$$A_{mn} = \left[\frac{n}{2^m}, \frac{n+1}{2^m}\right).$$

Observamos que para un m fijo, la familia $\{A_{nm}\}_{n\in\mathbb{Z}}$ es una familia ajena. Más aun, como $A_{nm} \subseteq \mathbb{R}$ es un conjunto medible, por el Lema 5 se sigue que la imagen inversa de cada A_{nm} bajo ϕ es un conjunto T-invariante.

$$\phi^{-1}(A_{mn}) = \left\{x \in X : \frac{n}{2^m} \leq \phi(x) < \frac{n+1}{2^m}\right\}.$$

Por hipótesis, T es una transformación ergódica, lo que implica que para cada $n \in \mathbb{Z}$, se tiene que $\mu(\phi^{-1}(A_{mn})) = 0$, o bien $\mu(X \setminus \phi^{-1}(A_{mn})) = 0$. La unión ajena de los conjuntos A_{mn} con m fija, nos da el espacio X, pues,

$$\bigcup_{n\in\mathbb{Z}} \phi^{-1}(A_{mn}) = \phi^{-1}\left(\bigcup_{n\in\mathbb{Z}}(A_{mn})\right) = \phi^{-1}(\mathbb{R}) = X.$$

Esto implica que para cada $m \in \mathbb{N}$, existe un único $n_m \in \mathbb{Z}$ tal que

$$\mu(X \setminus \phi^{-1}(A_{mn_m})) = 0.$$

Más aun, se tiene la siguiente sucesión acotada de conjuntos,

$$\cdots \subseteq A_{3n_3} \subseteq A_{2n_2} \subseteq A_{1n_1}.$$

De manera que la sucesión $(\frac{n_m}{2^m})_{m\in\mathbb{N}}$ es una sucesión creciente y acotada. Así pues, el límite existe. Ahora bien, definimos,

$$A = \bigcap_{m\in\mathbb{N}} A_{mn_m}.$$

Entonces,

$$\begin{aligned}\mu(X \setminus \phi^{-1}(A)) &= \mu\left(X \setminus \bigcap_{m\in\mathbb{N}} \phi^{-1}(A_{mn_m})\right) \\ &= \mu\left(\bigcup_{m\in\mathbb{N}} (X \setminus \phi^{-1}(A_{mn_m}))\right) \\ &= 0.\end{aligned}$$

Así pues, $A \neq \emptyset$. Sea $x \in A$, entonces para cada número natural m se tiene que,

$$0 \leq \left|\phi(x) - \frac{n_m}{2^m}\right| \leq \frac{1}{2^m}.$$

Esto implica que

$$\phi(x) = \lim_{m\to\infty} \frac{n_m}{2^m}.$$

Por lo tanto, la función ϕ es constante sobre A, o bien, ϕ es constante c.d. (rel μ).

Para probar el recíproco, supongamos ahora que todas las funciones medibles que son T-invariantes c.d. (rel μ) son constantes c.d. (rel μ). Sea $A \subseteq X$ un conjunto medible y T-invariante. Entonces, en particular la función característica χ_A es también medible y T-invariante, pues

$$\chi_A(T(x)) = \chi_{T^{-1}(A)}(x) = \chi_A(x) \qquad \forall x \in X.$$

Así pues, tenemos

$$\chi_A = 0 \qquad c.d.(rel \mu),$$

o bien,

$$\chi_A = 1 \qquad c.d.(rel \mu).$$

Por lo tanto, $\mu(A) =$ o $\mu(X \setminus A) = 0$, es decir, T es ergódica. □

Finalmente, veremos que en el caso en que el espacio X es de medida finita, o bien, $\mu(X) = 1$, podemos establecer que una transformación T es ergódica si y sólo si cada función integrable y T-invariante es también constante.

Teorema 20. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) < \infty$ y $T : X \to X$ una transformación medible. La transformación T es ergódica si y sólo si cada $\phi \in L^1(X, \mu)$ que es T-invariante c.d. (rel μ), resulta ser constante c.d. (rel μ).*

Demostración. Supongamos que T es una transformación ergódica. Sea ϕ una función integrable y T-invariante c.d. (rel μ). Veamos que es constante c.d. (rel μ). Para cada $r \in \mathbb{R}$, definimos el conjunto

$$E_r = \{x \in X : \phi(x) > r\}.$$

Observamos que dado que ϕ es T-invariante, se sigue que E_r es un conjunto medible y T-invariante. Como T es una transformación ergódica, se tiene también que $\mu(E_r) = 0$ o bien, $\mu(X \setminus E_r) = 0$. Si ϕ no fuera constante, se podría encontrar $r_0 \in \mathbb{R}$ tal que

$$0 < \mu(E_{r_0}) < \mu(X) < \infty.$$

Por lo tanto, ϕ debe ser constante c.d. (rel μ). Para probar el recíproco, supongamos que para cada función ϕ integrable y T-invariante c.d. (rel μ) se tiene que es constante c.d. (rel μ). Sea $A \subseteq X$ un conjunto medible y T-invariante c.d. (rel μ). Entonces se tiene que su función característica χ_A es una función medible y T-invariante. Más aun, se sabe que las funciones características de conjuntos medibles pertenecen al espacio $L^1(X, \mu)$. Por hipótesis se tiene que χ_A es constante c.d. (rel μ). Por lo tanto, $\mu(A) = 0$, o bien, $\mu(X \setminus A) = 0$. □

El siguiente teorema relaciona la noción de ergodicidad con el Teorema Ergódico de Birkhoff. Al inicio de la sección mencionamos que el concepto de ergodicidad tenía su origen en lo que se denominó la hipótesis ergódica. A continuación probaremos que si se supone que la transformación T es ergódica, entonces se satisface la hipótesis ergódica.

Teorema 21. *Sea* (X, S, μ) *un espacio de medida tal que* $\mu(X) < \infty$ *y sea* $T : X \to X$ *una transformación ergódica que preserva la medida* μ. *Si* $\phi \in L^1(X, \mu)$ *entonces,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)) = \frac{1}{\mu(X)} \int \phi \mathrm{d}\mu. \tag{3.1}$$

Para casi toda $x \in X$.

Demostración. Por el Teorema 18 se sigue que existe $\phi_T \in L^1(X, \mu)$ que es T-invariante c.d. (rel μ) tal que

$$\phi_T(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)).$$

Por el Teorema 20 se infiere que ϕ_T es constante c.d. (rel μ). Así también, por el Teorema 18 se tiene que

$$\int \phi_T \mathrm{d}\mu = \int \phi \mathrm{d}\mu.$$

Pero

$$\begin{aligned}
\int \phi_T \mathrm{d}\mu &= \int \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)) \mathrm{d}\mu \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)) \int \mathrm{d}\mu \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)) \mu(X).
\end{aligned}$$

Por lo tanto,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)) = \frac{1}{\mu(X)} \int \phi \mathrm{d}\mu.$$

□

En la siguiente sección veremos algunos ejemplos de transformaciones ergódicas. De hecho, analizaremos algunas de las transformaciones que preservan la medida que se estudiaron en el capítulo 1. Veremos en cada caso si la transformación es ergódica o al menos bajo qué condiciones sería ergódica.

3.2. Ejemplos de Transformaciones Ergódicas

A continuación se presenta una serie de ejemplos para ilustrar el concepto de ergodicidad. Los ejemplos que se muestran, se retoman del capítulo 1 y se prueba bajo qué condiciones son transformaciones ergódicas. Para probar la ergodicidad de la mayoría de los ejemplos, utilizamos su expansión en series de Fourier de las funciones integrables.

3.2.1. Transformación Expansión Binaria

A continuación probaremos que la transformación $T(x) = 2x \mod 1$ es una transformación ergódica y con base en el Teorema Ergódico de Birkhoff, calcularemos la frecuencia de los dígitos 0 y 1 en la expansión binaria de un número en el intervalo $(0, 1)$.

Proposición 18. *La transformación* $T(x) = 2x \mod 1$ *es ergódica.*

Demostración. Sea $f \in L^1(X, \mu)$ una función T-invariante. Veamos que f es constante. Consideremos los coeficientes de Fourier asociados a la función f,

$$a_n(f) = \int e^{-2\pi ikx} f(x)\mathrm{d}m.$$

Consideremos también los coeficientes de Fourier asociados a la función $f \circ T$,

$$\begin{aligned} a_n(f \circ T) &= \int e^{-2\pi ikT(x)} f(T(x))\mathrm{d}m \\ &= \int e^{-2\pi ik(2x)} f(2x))\mathrm{d}m. \end{aligned}$$

Haciendo un cambio de variable $y = 2x$ se tiene lo siguiente,

$$a_n(f \circ T) = \frac{1}{2} \int e^{-2\pi iky} f(y)\mathrm{d}m.$$

Así pues, $a_n(f \circ T) = \frac{1}{2} a_n(f)$ y entonces $a_n(f) = 0$ para cada n. Por lo tanto, f es constante. Por el Teorema 20 se tiene que T es ergódica. □

En el Ejemplo 9 mostramos como se puede escribir la expansión binaria de un número con ayuda de la transformación $T(x) = 2x \mod 1$. Posteriormente, con ayuda del Teorema de Recurrencia de Poincaré probamos que la expansión binaria de casi todo número en el intervalo $(0, 1)$ contiene una cantidad infinita de 0's y 1's. Para saber la frecuencia con la que aparecen en la expansión binaria, ocuparemos el Teorema Ergódico de Birkhoff y la ergodicidad de la Transformación Expansión Binaria. El teorema que se presenta a continuación recibe el nombre de *Teorema de Borel de los Números Normales.*

Teorema 22 (Borel de los Números Normales)**.** *La frecuencia en la que aparece el dígito 1 en la expansión binaria de todos los números en el intervalo* $(0, 1)$ *(salvo un conjunto de medida cero) es de* $\frac{1}{2}$.

Demostración. Sea $x \in (0, 1)$. Supongamos que su expansión binaria es la siguiente,

$$x = \frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots .$$

En el Ejemplo 9 mostramos que,

$$T(x) = T\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \frac{a_3}{2^3} + \cdots\right) = \frac{a_2}{2} + \frac{a_3}{2^2} + \frac{a_4}{2^3} + \cdots .$$

Sea $f(x) = \chi_{[\frac{1}{2},1)}(x)$. Entonces,

$$f(T^m(x)) = f\left(\frac{a_{m+1}}{2} + \frac{a_{m+2}}{2^2} + \cdots\right) = \begin{cases} 1 & \text{si y sólo si} \quad a_{m+1} = 1 \\ 0 & \text{si y sólo si} \quad a_{m+1} = 0. \end{cases}$$

El número de 1's en los primeros n dígitos en la expansión binaria de x está dado por,

$$\sum_{k=0}^{n-1} f(T^k(x)).$$

Dividiendo entre n, aplicando el Teorema Ergódico de Birkhoff y utilizando la ergodicidad de T se tiene,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \int \chi_{[\frac{1}{2},1)} \mathrm{d}m = \frac{1}{2}.$$

Por lo tanto, la frecuencia de 1 en la expansión binaria de un número en el intervalo $(0,1)$ es de $\frac{1}{2}$. $\square$

3.2.2. Rotaciones en el Círculo

En esta sección analizaremos las condiciones bajo las cuales las rotaciones en el círculo R_w definidas en la sección 1.2.2 son ergódicas. Una de las principales propiedades que tienen las rotaciones es que están relacionadas con las traslaciones T_α mediante el isomorfísmo $h(\theta) = e^{2\pi i\theta}$. Así pues, primero probaremos que las traslaciones racionales no son ergódicas, mientras que las traslaciones irracionales son ergódicas. A partir del isomorfísmo h, concluiremos que la rotación R_w es ergódica, sólo en el caso en que w sea de la forma $e^{2\pi i\alpha}$ con α irracional.

Hay dos maneras de ver que la transformación T_α no es ergódica. Una manera sería encontrar un conjunto T_α-invariante cuya medida fuera mayor que cero y menor que uno. Otra manera sería encontrar una función integrable que fuera T_α-invariante, pero no constante. A continuación examinaremos ambos caminos.

Sea $\alpha = \frac{1}{4}$ y T_α la traslación correspondiente. Consideremos el siguiente conjunto medible,

$$A = \left[0, \frac{1}{8}\right) \cup \left[\frac{1}{4}, \frac{3}{8}\right) \cup \left[\frac{1}{2}, \frac{5}{8}\right) \cup \left[\frac{4}{3}, \frac{7}{8}\right).$$

Observamos que el conjunto A es un conjunto T_α-invariante.

Sin embargo, $\mu(A) = \frac{1}{2}$, por lo que T_α no es una transformación ergódica. Así también, podemos encontrar funciones en el espacio $L^1(X,\mu)$ que son T_α-invariantes, pero que no son constantes. Por ejemplo, la función $f(x) =$

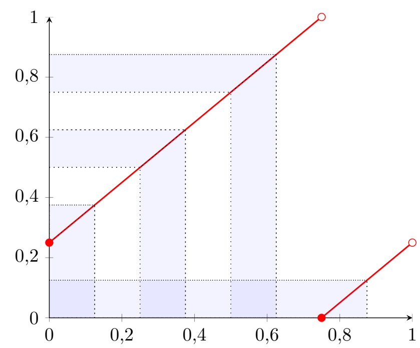


Figura 3.2: A es un conjunto T_α-invariante.

$\cos(2\pi 4x)$. Observamos que es T_α invariante. En efecto,

$$\begin{aligned} f(T_\alpha(x)) &= \cos(2\pi 4(x+\frac{1}{4})) \\ &= \cos(2\pi 4x + 2\pi) \\ &= \cos(2\pi 4x)\cos(2\pi) - \sin(2\pi 4x)\sin(2\pi) \\ &= \cos(2\pi 4x) \\ &= f(x). \end{aligned}$$

Sin embargo, f no es constante. Por lo tanto, T_α no es ergódica. El argumento se puede generalizar para cualquier valor racional de α.

Proposición 19. *Sea $\alpha \in (0,1)$. Si α es irracional, entonces la traslación T_α es ergódica.*

Demostración. Sea $f \in L^1(X,\mu)$ una función T_α-invariante. Se verá que f es constante. Consideremos la serie de Fourier asociada a f,

$$f(x) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}.$$

Consideremos también la serie de Fourier asociada a $f \circ T_\alpha$,

$$\begin{aligned} f(T_\alpha(x)) &= \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n T_\alpha(x)} \\ &= \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n(x+\alpha)} \\ &= \sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x} e^{2\pi i n\alpha}. \end{aligned}$$

Como f es T-invariante, tenemos que $f(x) = f(T(x))$. Entonces,

$$f(x) - f(T(x)) = \sum_{n\in\mathbb{Z}} a_n(1 - e^{2\pi i n\alpha})e^{2\pi i n x} = 0.$$

Por la unicidad de los coeficientes de Fourier,

$$a_n(1 - e^{2\pi i n\alpha}) = 0 \qquad \forall n \in \mathbb{Z}.$$

Como por hipótesis α es irracional, entonces $1 - e^{2\pi i n\alpha} \neq 0$. Por lo tanto,

$$a_n = 0 \qquad \forall n \neq 0.$$

Concluimos que $f(x) = a_0$ para toda $x \in X$, es decir, f es constante y por el Teorema 20 se tiene que T es ergódica. $\square$

A continuación probaremos que si se toma $w = e^{2\pi i\alpha}$ con α irracional, entonces la rotación R_w es ergódica. En la Sección 1.2.2 vimos que las rotaciones y las traslaciones se relacionaban mediante el isomorfísmo $h(\theta) = e^{2\pi i\theta}$, de manera que $R_w \circ h = h \circ T_\alpha$.

Proposición 20. *Sea $w \in \mathbb{C}$ definido como $w = e^{2\pi i\alpha}$ con $\alpha \in (0,1)$ irracional. La rotación R_w es ergódica.*

Demostración. Sea $f \in L^1(X,\mu)$ una función R_w-invariante. Veamos que f es constante. Como sabemos que $R_w \circ h = h \circ T_\alpha$, entonces $R_w = h \circ T_\alpha \circ h^{-1}$.

$$\begin{aligned} f(x) &= f(R_w(x)) \\ &= f((h \circ T_\alpha \circ h^{-1})(x)) \\ &= f \circ h \circ T_\alpha \circ h^{-1}(x). \end{aligned}$$

Es decir, $f = f \circ h \circ T_\alpha \circ h^{-1}$. Entonces $f \circ h = f \circ h \circ T_\alpha$. La función $f \circ h \in L^1(X,\mu)$ es T_α-invariante. Por la proposición anterior, se sabe que T_α es una transformación ergódica y por el Teorema 20 se sigue que $f \circ h$ debe ser constante c.d. (rel μ). Supongamos que $f \circ h = c_0$, donde $c_0 \in \mathbb{R}$. Esto implica que, $f = c_0 \circ h^{-1} = c_0$ c.d. (rel μ). Por el Teorema 20, se concluye que R_w es una transformación ergódica. $\square$

3.2.3. Automorfismos en el Toro

En esta sección examinaremos las condiciones bajo las cuales los automorfismos en el toro son transformaciones ergódicas. Los automorfismos en el toro son ergódicos cuando al considerar cada iteración de la matriz que los induce, encontramos que ninguna iteración tiene como valor característico al uno.

Proposición 21. *Sea $A \in S(k,\mathbb{Z})$. Un automorfismo T_A en el toro n'ésimo es ergódico si y sólo si ningún valor característico de A es una raíz de la unidad.*

Demostración. Sea $f \in L^1(X,\mu)$ una función T_A-invariante. Por hipótesis $A \in S(k,\mathbb{Z})$, esto implica que $|\det A| = 1$. Así pues, para cada $m \in \mathbb{Z}^k$ consideramos los coeficientes de Fourier de $f \circ T_A$. Haciendo un cambio de variable obtenemos,

$$\begin{aligned} a_m(f \circ T_A) &= \int e^{-2\pi i\langle m,x\rangle} f(T_A(x))\mathrm{d}\lambda \\ &= \int e^{-2\pi i\langle m,x\rangle} f(A(x))\mathrm{d}\lambda \\ &= \int e^{-2\pi i\langle m,A^{-1}x\rangle} f(x)\mathrm{d}\lambda \\ &= \int e^{-2\pi i\langle (A^{-1})^* m,x\rangle} f(x)\mathrm{d}\lambda \\ &= a_{(A^{-1})^* m}(f). \end{aligned}$$

Como supusimos que f es T_A-invariante, entonces se tiene que $a_m(f \circ T_A) = a_m(f)$.

$$a_m(f) = a_{(A^{-1})^* m}(f) \qquad \forall m \in \mathbb{Z}^k.$$

Observamos que (A^*) no tiene raíces de la unidad como valores característicos, es decir, 1 no es un valor característico de (A^*), pues si lo fuera, implicaría que 1 también es un valor característico de A^m. Pero por hipótesis, A no tiene raíces de la unidad como valores característicos. Ahora bien, por el *Lema de Riemann-Lebesgue*, se sigue que,

$$a_{(A^*)^n m}(f) \to 0 \qquad \text{si } n \to \infty.$$

Por lo tanto, $a_m(f) = 0$ para cada $m \neq 0$. Así pues, $f(x) = a_0(f)$ c.d. (rel λ) y por el Teorema 20 se tiene que T_A es ergódica. $\square$

3.2.4. Cadenas de Markov

En la sección 1.2.5 definimos los espacios Σ_k y Σ_k^+ de sucesiones bilaterales y laterales respectivamente. Posteriormente trabajamos con algunos subespacios Σ_A y Σ_A^+ llamados cadenas de Markov bilaterales y laterales respectivamente. Para dichos espacios definimos la transformación corrimiento σ. Se probó que la transformación σ preservaba tanto la medida de Markov como la medida de Bernoulli. En esta sección veremos que la transformación corrimiento es una transformación ergódica con respecto a la medida de Bernoulli y que es una transformación ergódica con respecto a la medida de Markov cuando la matriz estocástica es irreducible.

Medidas de Bernoulli.

El argumento que presentamos para probar que la transformación corrimiento σ es una transformación ergódica con respecto a la medida de Bernoulli funciona tanto para los corrimientos laterales, como bilaterales. Así pues, no haremos distinción alguna. Antes de presentar el resultado principal de la sección, probaremos tres lemas auxiliares de Teoría de la Medida. El primer lema afirma que todo conjunto medible puede ser aproximado tanto como se quiera por elementos del álgebra de conjuntos que genera la σ-álgebra.

Lema 6. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Si $S = S(\mathcal{A})$, donde $\mathcal{A}$ es un álgebra que genera la σ-álgebra S, entonces para cada $E \subseteq X$ medible y para cada $\epsilon > 0$, existe $A \in \mathcal{A}$ tal que $\mu(A \triangle E) < \epsilon$.*

Demostración. Sea $\epsilon > 0$ y $E \subseteq X$ un conjunto medible. Por definición, se sabe que

$$\mu(E) = \text{ínf}\left\{\sum_{n=1}^{\infty} \mu(A_n) : A_n \text{ es cubierta de } E \text{ y } (A_n)_{n=1}^{\infty} \subseteq \mathcal{A}\right\}.$$

Por definición de ínfimo, hay $(A_i)_{i\in\mathbb{N}}$ una cubierta de E tal que,

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu(A_i) < \mu(E) + \frac{\epsilon}{2}.$$

Por hipótesis, $\mu(E) < \infty$ y entonces $\mu(E) + \frac{\epsilon}{2} < \infty$. Tomamos $A = \cup_{n=1}^{\infty} A_n \in S$. Entonces

$$\mu(A) \leq \sum_{i=1}^{\infty} \mu(A) < \infty.$$

Proponemos $F_n = \bigcup_{i=1}^{n} A_i \in \mathcal{A}$ para cada $n \in \mathbb{N}$. Observamos que

$$F_n \subseteq F_{n+1} \qquad \forall n \in \mathbb{N}.$$

También observamos que

$$\bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} A_n = A.$$

Así pues,

$$\mu(A) = \lim_{n\to\infty} \mu(F_n).$$

Para $\epsilon > 0$ dada, existe $N \in \mathbb{N}$ tal que

$$|\mu(F_n) - \mu(A)| < \epsilon \qquad \forall n \geq N.$$

En particular, $F_N = A' \in \mathcal{A}$ y

$$\mu(A) - \mu(A') < \frac{\epsilon}{2}.$$

Ahora bien, veamos que $\mu(E \triangle A) < \epsilon$. Notamos que

$$\mu(E \triangle A') = \mu(E \setminus A') + \mu(A' \setminus E).$$

Como $E \subseteq \bigcup_{i=1}^{\infty} A_i = A$, entonces

$$\mu(A \setminus E) = \mu(A) - \mu(E) < \frac{\epsilon}{2}.$$

También como $E \setminus A' \subseteq A \setminus A'$, por monotonía, tenemos que $\mu(E \setminus A') \leq \mu(A' \setminus A') = \mu(A') - \mu(A) < \frac{\epsilon}{2}$. Por lo tanto,

$$\mu(E \triangle A') < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

$\square$

El siguiente lema nos da una cota al valor absoluto de la diferencia de las medidas de dos conjuntos dentro de un espacio de medida finita.

Lema 7. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Si $A, B \in S$, entonces $|\mu(A) - \mu(B)| \leq \mu(A \triangle B)$.*

Demostración. Observamos que $A = A \cap X = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$. Como $A \cap B \subseteq B$ y $A \cap B^c = A \setminus B \subseteq A \triangle B$, entonces tenemos

$$(A \cap B) \cup (A \cap B^c) \subseteq B \cup (A \triangle B).$$

Entonces

$$A \subseteq B \cup (A \triangle B).$$

Por monotonía y la σ-aditividad de la medida μ,

$$\mu(A \cap B) \leq \mu(B) + \mu(A \triangle B).$$

Como el espacio tiene medida finita,

$$\mu(A) - \mu(B) \leq \mu(A \triangle B).$$

De manera análoga obtenemos

$$\mu(B) - \mu(A) \leq \mu(A \triangle B).$$

Por lo tanto,

$$|\mu(A) - \mu(B)| \leq \mu(A \triangle B).$$

$\square$

El siguiente lema es técnico en el sentido de que nos auxilia en la prueba de la ergodicidad de σ.

Lema 8. *Sean $E, A \subseteq X$ y sea $T : X \to X$ una transformación. Entonces,*

$$E \triangle (A \cap T^{-1}(A)) \subseteq (E \triangle A) \cup (E \triangle T^{-1}(A)).$$

Demostración. Por contención inversa, supongamos que $x \notin (E \triangle A) \cup (E \triangle T^{-1}(A))$. Esto implica que $x \notin (E \triangle A)$ y que $x \notin (E \triangle T^{-1}(A))$. Pero

$$x \notin (E \triangle A) \Leftrightarrow \chi_E(x) = \chi_A(x).$$

También,

$$x \notin (E \triangle T^{-1}(A)) \Leftrightarrow \chi_E(x) = \chi_{T^{-1}(A)}(x).$$

Así pues se sigue

$$\chi_E(x) = (\chi_E(x))^2 = \chi_A(x)\chi_{T^{-1}(A)}(x) = \chi_{A\cap T^{-1}(A)}(x),$$

pero esto ocurre si y sólo si $x \notin E\triangle(A\cap T^1(A))$. Por lo tanto, $E\triangle(A\cap T^{-1}(A)) \subseteq (E\triangle A) \cup (E\triangle T^{-1}(A))$. □

A continuación se prueba que la transformación corrimiento σ es una transformación ergódica. El argumento que se presenta funcionaba tanto para sucesiones laterales como para sucesiones bilaterales. Así pues, se puede suponer que el espacio con el que se trabaja es el espacio Σ_k o bien, el espacio Σ_k^+ con la σ-álgebra generada por los respectivos cilindros.

Proposición 22. *σ es una transformación ergódica con respecto a la medida de Bernoulli μ.*

Demostración. Sea $E \subseteq \Sigma_k$ un conjunto medible y σ-invariante. Se verá que $\mu(E) = 0$ o bien, $\mu(E) = 1$. Sea $\epsilon' > 0$, tomamos $\epsilon = \frac{\epsilon'}{\text{máx}\{\mu(E),\mu(A)\}}$ por el Lema 6 se tiene que existe A que es la unión finita de cilindros tal que

$$\mu(E\triangle A) < \epsilon.$$

Por el Lema 7 se tiene,

$$|\mu(E) - \mu(A)| \leq \mu(E\triangle A) < \epsilon.$$

Como A es una unión finita de cilindros, entonces depende de un número finito de coordenadas, luego existe $n_0 \in \mathbb{N}$ tal que el conjunto $\sigma^{-n_0}(A)$ es un cilindro que depende de coordenadas diferentes que A. Como μ es la medida de Bernoulli y la transformación corrimiento σ preserva la medida μ se tiene lo siguiente,

$$\begin{aligned} \mu(A \cap \sigma^{-n_0}(A)) &= \mu(A)\mu(\sigma^{-n_0}(A)) \\ &= \mu(A)\mu(A) = \mu(A)^2. \end{aligned}$$

Más aún, como E es σ-invariante y σ preserva la medida de Bernoulli,

$$\begin{aligned} \mu(E\triangle\sigma^{-n_0}(A)) &= \mu(\sigma^{-n_0}(E)\triangle\sigma^{-n_0}(A)) \\ &= \mu(\sigma^{-n_0}(E\triangle A)) \\ &= \mu(E\triangle A) < \epsilon. \end{aligned}$$

Por el Lema 8 se tiene

$$E\triangle(A \cap \sigma^{-n_0}(A)) \subseteq (E\triangle A) \cup (E\triangle\sigma^{-n_0}(A)).$$

Esto implica que, dado que la unión en disjunta por la elección de n_0,

$$\mu(E\triangle(A \cap \sigma^{-n_0}(A))) \leq \mu(E\triangle A) + \mu(E\triangle\sigma^{-n_0}(A)) < 2\epsilon.$$

Así pues,

$$|\mu(E) - \mu(A \cap \sigma^{-n_0}(A))| \leq \mu(E\triangle(A \cap \sigma^{-n_0}(A))) < 2\epsilon.$$

Por la desigualdad del triángulo y las ecuaciones anteriores se tiene

$$\begin{aligned}|\mu(E)-\mu(E)^2| &= |\mu(E)-\mu(A)^2| + |\mu(A)^2-\mu(E)|\\ &= |\mu(E)-\mu(A\cap\sigma^{-n_0}A)| + (\mu(A)+\mu(E))|\mu(A)-\mu(E)|\\ &= 2\epsilon + \mu(A)|\mu(A)-\mu(E)| + \mu(E)|\mu(A)-\mu(E)|\\ &< 4\epsilon'.\end{aligned}$$

Como ϵ' es arbitrario, se sigue que $\mu(E)=\mu(E)^2$. Sin embargo, esto sólo puede ocurrir en caso de que $\mu(E)=0$ o bien, $\mu(E)=1$. Por lo tanto, σ es una transformación ergódica. $\square$

Medidas de Markov.

La transformación corrimiento σ con respecto a la medida de Markov no siempre es ergódica. En esta sección se probarán las condiciones bajo las cuales σ es ergódica con respecto a la medida de Markov. Al igual que en la sección anterior, el argumento que se presenta funciona tanto para los corrimientos laterales como los bilaterales, de manera que no haremos distinción alguna.

Supongamos que $P=(p_{ij})$ es una matriz estocástica. Recordamos que P es una matriz estocástica si es una matriz cuadrada de $k\times k$, sus entradas son no negativas y cumple la suma de sus filas es igual a 1. Definimos,

$$Q = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}P^k. \tag{3.2}$$

Donde $P^k=(p_{ij}^{(k)})$ es la k'ésima potencia de la matriz P y $P^0=Id$. Denotamos $Q=(q_{ij})$ donde,

$$q_{ij} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}p_{ij}^{(k)}.$$

La siguiente proposición muestra que Q está bien definida.

Proposición 23. *Para cada $i,j\in\{0,1,\cdots N-1\}$, el límite*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}p_{ij}^{(k)},$$

existe.

Demostración. Consideramos los cilindros C_i y C_j con $i,j\in\mathcal{A}$. Entonces,

$$\sigma^{-n}(C_i)\cap C_j = \bigcup_{a_1,\cdots,a_{n-1}} C_{i,a_1,\cdots,a_{n-1},j}.$$

Entonces,

$$\begin{aligned}\mu(\sigma^{-n}(C_i)\cap C_j) &= \sum_{a_1,\cdots,a_{n-1}} \mu(C_{i,a_1,\cdots,a_{n-1},j})\\ &= \sum_{a_1,\cdots,a_{n-1}} p_i p_{ia_1}\cdots p_{a_{n-1}j}\\ &= p_i(P^n)_{ij}.\end{aligned}$$

Esto implica que,

$$q_{ij} = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}(p^n)_{ij} = \lim_{n\to\infty}\frac{1}{p_i}\frac{1}{n}\sum_{k=0}^{n-1}\mu(\sigma^{-n}(C_i)\cap C_j).$$

Como $\chi_{C_i}\in L(X,\mu)$, por el Teorema Ergódico de Birkhoff se tiene que,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\chi_{C_i}(\sigma^k(x)) = \phi_\sigma(x) \qquad c.d.(rel\mu).$$

Donde $\phi_\sigma(x)$ satisface,

$$\int \phi_\sigma \mathrm{d}\mu = \int \chi_{C_i}\mathrm{d}\mu = \mu(C_i) = p_i.$$

Multiplicamos por χ_{C_j} ambos lados de la igualdad.

$$\begin{aligned}\chi_{C_j}(x)\phi_\sigma(x) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\chi_{C_j}(x)\chi_{C_i}(\sigma^k(x)) \qquad c.d.(rel\mu)\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\chi_{C_j\cap\sigma^{-1}(C_i)}(x) \qquad c.d.(rel\mu).\end{aligned}$$

Integrando y por el Teorema de la Convergencia Monótona,

$$\begin{aligned}\int \chi_{C_j}\phi_\sigma\mathrm{d}\mu &= \int \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\chi_{C_j\cap\sigma^{-1}(C_i)}(x)\mathrm{d}\mu\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\int\chi_{C_j\cap\sigma^{-1}(C_i)}(x)\mathrm{d}\mu\\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(C_j\cap\sigma^{-1}(C_i))\\ &= p_i q_{ij}.\end{aligned}$$

Así pues,

$$\begin{aligned} q_{ij} &= \frac{1}{p_i}\int \chi_{C_j}\phi_\sigma \mathrm{d}\mu \\ &= \frac{1}{p_i}\int_{C_j}\phi_\sigma \mathrm{d}\mu \\ &\leq \frac{1}{p_i}\int \phi_\sigma \mathrm{d}\mu \\ &= \frac{1}{p_i}p_i = 1. \end{aligned}$$

Por lo tanto q_{ij} está bien definida. □

El producto de dos matrices estocásticas A, B es a su vez una matriz estocástica. En efecto,

$$\sum_{j=1}^{k}(AB)_{ij} = \sum_{j=1}^{k}\sum_{l=1}^{k}A_{il}B_{lj} = \sum_{l=1}^{k}A_{il}\sum_{j=1}^{k}B_{lj} = \sum_{l=1}^{k}A_{il} = 1.$$

Así pues P^n es una matriz estocástica para cada $n \in \mathbb{N}$ y también,

$$\frac{1}{n}\sum_{k=0}^{n-1}P^k.$$

Como,

$$\begin{aligned} \sum_{j=1}^{k} q_{ij} &= \sum_{j=1}^{k}\lim_{n\to\infty}\frac{1}{n}\sum_{l=0}^{n-1}p_{ij}^{(l)} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{=0}^{n-1}\sum_{j=1}^{k}p_{ij}^{(l)} \\ &= \lim_{n\to\infty}\frac{1}{n}\sum_{=0}^{n-1}1 \\ &= 1. \end{aligned}$$

Por lo tanto, Q es una matriz estocástica. El siguiente lema prueba algunas propiedades de la matriz Q.

Lema 9. *Sea P una matriz estocástica, Q definida como en (3.2) y v un vector. Se satisfacen las siguientes propiedades:*

1. *$QP = PQ = Q$.*
2. *Si $vP = v$, entonces $vQ = v$.*
3. *$Q^2 = Q$.*

Demostración. 1) Sea $n \in \mathbb{N}$. Consideramos la matriz estocástica $\frac{1}{n}\sum_{k=0}^{n-1} P^k$ y la multiplicamos por P,

$$\left(\frac{1}{n}\sum_{k=1}^{n} P^k\right) P = P\left(\frac{1}{n}\sum_{k=1}^{n} P^k\right) = \left(\frac{1}{n}\sum_{k=1}^{n+1} P^k\right).$$

Haciendo tender n al infinito obtenemos,

$$\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} P^k\right) P = P\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} P^k\right) = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n+1} P^k\right).$$

Por definición de Q,

$$QP = PQ = Q.$$

2) Sea $n \in \mathbb{N}$. Supongamos que se cumple $vP = v$, entonces también se cumple que $vP^n = v$ para cada n natural. Así también,

$$v\left(\frac{1}{n}\sum_{k=1}^{n} P^k\right) = v.$$

Tomando límites y por definición de Q obtenemos,

$$vQ = v.$$

3) Por propiedad 1) que acabamos de probar, sabemos que $QP = Q$, y entonces tenemos que $QP^n = Q$ para cada $n \in \mathbb{N}$. También,

$$Q\left(\frac{1}{n}\sum_{k=1}^{n} P^k\right) = Q.$$

Tomando límites concluimos, $Q^2 = Q$. □

El siguiente teorema prueba una equivalencia de la ergodicidad que se utilizará al probar las condiciones bajo las cuales una cadena de Markov es ergódica.

Teorema 23. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Sea $T : X \to X$ una transformación que preserva la medida μ. La transformación T es ergódica si y sólo si para cualesquiera $A, B \subseteq X$ medibles se satisface,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(T^{-k}(A)\cap B) = \mu(A)\mu(B).$$

Demostración. Supongamos primero que T es ergódica. Sean $A, B \subseteq X$ conjuntos medibles y consideramos la función característica $\chi_A \in L^1(X,\mu)$. Por el Teorema 21 tenemos,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\chi_A(T^k(x)) &= \int \chi_A \mathrm{d}\mu \\ &= \mu(A) \qquad c.d.\mathrm{rel}(\mu).\end{aligned}$$

Por el Teorema de la Convergencia Monótona,

$$\begin{aligned}
\mu(A)\mu(B) &= \mu(A)\int \chi_B \mathrm{d}\mu \\
&= \int \mu(A)\chi_B \mathrm{d}\mu \\
&= \int \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \chi_A(T^k(x))\chi_B \mathrm{d}\mu \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int \chi_A(T^k(x))\chi_B \mathrm{d}\mu \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \int \chi_{T^{-k}(A)\cap B} \mathrm{d}\mu \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(A)\cap B).
\end{aligned}$$

Recíprocamente, supongamos ahora que $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(A)\cap B) = \mu(A)\mu(B)$ para cualesquiera A, B conjuntos medibles y veamos que la transformación T es ergódica. Sea A un conjunto medible y T-invariante. Entonces tenemos lo siguiente,

$$\begin{aligned}
\mu(A)\mu(X\setminus A) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(T^{-k}(A)\cap X\setminus A) \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(A\cap X\setminus A) \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mu(\emptyset) \\
&= 0.
\end{aligned}$$

Así pues, $\mu(A) = 0$, o bien, $\mu(X\setminus A) = 0$. Por lo tanto, T es ergódica. □

A continuación se prueban cinco equivalencias de que un corrimiento σ es ergódico con respecto a la medida de Markov.

Proposición 24. *Los siguientes enunciados son equivalentes,*

1. *σ es ergódica.*
2. *Las columnas de Q son idénticas.*
3. *$q_{ij} > 0$ para todo i, j.*
4. *P es irreducible.*

5. *1 es un valor característico simple de P.*

Demostración.
1) $\Rightarrow$ 2)

Supongamos que σ es ergódica. Consideramos los cilindros C_i, C_j y sus funciones características $\chi_{C_i}, \chi_{C_j} \in L^1(X, \mu)$. Por el Teorema 21 tenemos,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{C_j}(\sigma^k(x)) = \int \chi_{C_j} \mathrm{d}\mu = \mu(C_j) = p_j.$$

Multiplicamos la igualdad por χ_{C_i},

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{C_i \cap \sigma^{-k}(C_j)}(x) = \chi_{C_i} p_j.$$

Integrando y por el Teorema de la Convergencia Dominada,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(C_i \cap \sigma^{-k}(C_j)) = p_i p_j.$$

Así pues,

$$q_{ij} = \frac{1}{p_i} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(C_i \cap \sigma^{-k}(C_j)) = p_j.$$

Por lo tanto, $q_{ij} = p_j$, es decir, todas las columnas de Q son idénticas.
2) $\Rightarrow$ 3)

Supongamos que todas las columnas de Q son idénticas. Entonces para cada j existe una constante c_j tal que $q_{ij} = c_j$. Por el Lema 9 se tiene que como p es tal que $pP = p$, entonces $pQ = p$ y sabemos que $p_j > 0$ para toda $j \in \{1, 2, \cdots, N-1\}$. Así pues,

$$p_j = \sum_{i=1}^{N-1} p_i q_{ij} = \sum_{i=1}^{N-1} p_i c_j = c_j \sum_{i=1}^{N-1} p_i = c_j = q_{ij}$$

para toda i, j.
3) $\Rightarrow$ 4)

Supongamos que $q_{ij} > 0$ para cada $i, j \in \{1, 2, \cdots, N-1\}$. Observamos que para todo i, j se tiene que,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (P^{(k)})_{ij} = q_{ij} > 0.$$

Entonces existe $n \geq 1$ tal que $(P^{(n)})_{ij} > 0$. Por lo tanto, la matriz P es irreducible.

4) $\Rightarrow$ 3)

Supongamos que P es irreducible. Veamos que los elementos de la matriz Q son positivos. Para cada $i \in \{0, 1, \cdots, N-1\}$ definimos el conjunto de los elementos positivos de la fila i,

$$S_i = \{j : q_{ij} > 0\}.$$

Como Q es una matriz estocástica, se tiene que $S_i \neq \emptyset$, pues sus filas deben sumar 1. Sea $l \in S_i$. Entonces $q_{il} > 0$. Como P es irreducible, existe un n natural tal que $(p^{(n)})_{lj} > 0$. Por el Lema 9 tenemos que $Q = QP = QP^n$. Así pues, para cada j se obtiene,

$$q_{ij} = \sum_{m=0}^{N-1} q_{im}(p^{(n)})_{mj} \geq q_{il}(p^{(n)})_{lj} > 0.$$

Por lo tanto, $q_{ij} > 0$ para cada $i, j \in \{1, 2, \cdots, N-1\}$.

3) $\Rightarrow$ 2)

Supongamos que $q_{ij} > 0$ para cada $i, j \in \{1, 2, \cdots N-1\}$. Fijamos un j. Por contradicción, supongamos que los q_{ij} son diferentes y definimos elemento más grande de la columna como

$$q_j = \max_{0 \leq i \leq N-1} q_{ij}.$$

Por el Lema 9 sabemos que Q es una matriz estocástica y que $Q^2 = Q$. Esto implica que para cada $i \in \{1, 2, \cdots N-1\}$,

$$q_{ij} = \sum_{l=1}^{N-1} q_{il}q_{lj} < \sum_{l=0}^{N-1} q_{il}q_j = q_j.$$

Por lo tanto, $q_{ij} < q_j$ para todo $i \in \{1, 2, \cdots N-1\}$, lo cual es una contradicción por la definición de q_j, pues al menos un elemento de la columna tendría que ser igual a q_j. Por lo tanto, las filas de Q deben ser idénticas.

2) $\Rightarrow$ 1)

Supongamos que las columnas de Q son idénticas, es decir, que $q_{ij} = p_j$ para todo $i, j \in \{1, 2, \cdots N-1\}$. Veamos que σ es ergódica. Sabemos que

$$p_j = q_{ij} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}.$$

Sean $C_{i_0, \cdots, i_l}, C_{j_0, \cdots, j_m}$ dos cilindros. Como σ es la transformación corrimiento, existe n natural lo suficientemente grande tal que $\sigma^{-n}(C_{i_0, \cdots, i_l})$ y $C_{j_0, \cdots, j_m}$

dependan de diferentes coordenadas. Así pues, para n suficientemente grande, tenemos,

$$\mu(\sigma^{-n}(C_{i_0,\cdots,i_l}) \cap C_{j_0,\cdots,j_m}) = p_{j_0}p_{j_0j_1}\cdots p_{j_{m-1}j_m}p_{j_mi_0}p_{i_0i_1}\cdots p_{i_{l-1}i_l}.$$

Así,

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mu(\sigma^{-k}(C_{i_0,\cdots,i_l}) \cap C_{j_0,\cdots,j_m}) \\
&\qquad = p_{j_0}p_{j_0j_1}\cdots p_{j_{m-1}j_m}p_{i_0i_1}\cdots p_{i_{l-1}i_l}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}p^{(k)}_{j_mi_0} \\
&\qquad = (p_{j_0}p_{j_0j_1}\cdots p_{j_{m-1}j_m})(p_{i_0}p_{i_0i_1}\cdots p_{i_{l-1}i_l}). \\
&\qquad = \mu(C_{j_0,\cdots,j_m})\mu(C_{i_0,\cdots,i_l}).
\end{aligned}$$

Por el Teorema 23 concluimos que σ es ergódica.

2) $\Rightarrow$ 5)

Supongamos que las columnas de Q son idénticas y veamos que 1 es un eigenvalor simple de P, es decir, supongamos que

$$q_{ij} = p_j \qquad \forall i,j \in \{1,2,\cdots,N-1\}.$$

Suponegamos que v es tal que $vP = v$. Por el Lema 9 sabemos que $vQ = v$. Esto implica

$$\begin{aligned}
v &= vQ \\
&= \left(p_1\sum_{i=1}^{N-1}v_i, p_2\sum_{i=1}^{N-1}v_i,\cdots,p_{N-1}\sum_{i=1}^{N-1}v_i,\cdots\right) \\
&= \left(\sum_{i=1}^{N-1}v_i\right)(p_1,p_2,\cdots,p_{N-1}) \\
&= \left(\sum_{i=1}^{N-1}v_i\right)p
\end{aligned}$$

Así pues, v es un múltiplo de p. Por lo tanto, 1 es una raíz característica simple de P.

5) $\Rightarrow$ 2)

Supongamos que 1 es una raíz característica simple. Veamos que las columnas de Q son idénticas. Consideremos las filas de la matriz Q

$$q^*_i = (q_{i0},\cdots,q_{i(N-1)}) \qquad \forall i \in \{1,2,\cdots N-1\}.$$

Por el Lema 9 se tiene que $Q = QP$. Esto implica que $q_i^* = q_i^* P$, pero por hipótesis, 1 es una raíz característica simple, esto implica que existe un único vector p tal que $p = pP$. Por lo tanto, $p = q_i^*$ y así, las filas de Q son idénticas. □

3.3. Ergodicidad Única

En la sección 1.3 probamos que el problema de la existencia de medidas invariantes siempre tiene solución, es decir, dado un espacio de medida (X, S, μ') y una transformación medible $T : X \to X$, siempre se puede encontrar una medida $\mu : S \to \mathbb{R}^+$ tal que T preserva la medida μ siempre y cuando el espacio X sea un espacio métrico compacto y T sea continua. Sin embargo, no se mencionó nada acerca de la unicidad de la medida μ. En esta sección veremos que aunque no siempre se puede garantizar la unicidad de una medida μ, cuando se pueda garantizar que es única, se puede fortalecer el Teorema Ergódico de Birkhoff.

En la sección 2.2 probamos en el Teorema Ergódico de Birkhoff que dado un espacio de medida (X, S, μ) y una transformación T que preserva la medida μ, si $\phi \in L^1(X, \mu)$, entonces para cada $x \in X$ el siguiente límite siempre existe,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)).$$

Ahora bien, cuando la medida μ que es preservada por T es única, se puede garantizar que la convergencia que asegura el Teorema Ergódico de Birkhoff es uniforme.

3.3.1. Ergodicidad y Convergencia Uniforme

A continuación definiremos la noción de ergodicidad única de manera formal y probaremos que la ergodicidad única en el Teorema Ergódico de Birkhoff garantiza la convergencia uniforme.

Definición 22. *Sea (X, S, μ) y sea $T : X \to X$ una transformación que preserva la medida μ. Decimos que la transformación T es* ***únicamente ergódica*** *si μ es la única medida preservada por la transformación T.*

La siguiente proposición muestra que las transformaciones únicamente ergódicas son transformaciones ergódicas.

Proposición 25. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Si T es una transformación únicamente ergódica, entonces T es una transformación ergódica.*

Demostración. Por hipótesis, T es una transformación únicamente ergódica. Por contrapositiva, supongamos que T no es ergódica, Entonces tenemos que existe $E \subseteq X$ medible y T-invariante tal que,

$$0 < \mu(E) < 1.$$

Definimos las funciones $\mu_1 : S \to \mathbb{R}$ y $\mu_2 : S \to \mathbb{R}$ como sigue, para cada $B \subseteq X$ medible,

$$\mu_1(B) = \tfrac{\mu(B\cap E)}{\mu(E)} \quad \mu_2(B) = \tfrac{\mu(B\cap(X\setminus E))}{\mu(X\setminus E)}.$$

Observamos que las funciones μ_1 y μ_2 son medidas porque μ es medida. Además, $\mu_1 \neq \mu_2$ pues,

$$\mu_1(E) = \frac{\mu(E\cap E)}{\mu(E)} = \frac{\mu(E)}{\mu(E)} = 1.$$

Mientras que,

$$\mu_2(E) = \frac{\mu(E\cap(X\setminus E))}{\mu(X\setminus E)} = \frac{\mu(\emptyset)}{\mu(E)} = 0.$$

Así pues, T no es únicamente ergódica. Por lo tanto, si T es únicamente ergódica, entonces tiene que ser ergódica. □

A continuación vamos a construir una medida que nos permitirá probar que la convergencia en el Teorema Ergódico de Birkhoff es uniforme en el caso en que la transformación T es únicamente ergódica. Definimos la función $\delta_x : S \to \mathbb{R}$ como sigue,

$$\delta_x(\{y\}) = \begin{cases} 1 & \text{si } y = x \\ 0 & \text{si } y \neq x. \end{cases}$$

La función δ_x es una medida, pues cumple,

1. $\delta_x(\emptyset) = 0$.
2. $\delta_x(E) \geq 0$ para todo $E \in S$.
3. Si $(E_n)_{n\in\mathbb{N}}$ es una sucesión de elementos disjuntos de S, entonces

$$\delta_x\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \delta_x(E_n).$$

En efecto, la primera propiedad se sigue directamente de la definición. Para verificar la segunda propiedad, consideremos $E \in S$. Pueden ocurrir dos casos:
Caso 1: $x \in E$, entonces,

$$\delta_x(E) = \delta_x(\{x\}) = 1 \geq 0.$$

Caso 2: $x \notin E$, entonces,

$$\delta_x(E) = \delta_x(\emptyset) = 0 \geq 0.$$

Finalmente, para verificar la tercera propiedad, consideremos una sucesión $(E_n)_{n\in\mathbb{N}}$ de elementos disjuntos de la σ-álgebra. Tenemos los siguientes casos:
Caso 1: $x \in \bigcup_{n=1}^{\infty} E_n$. Entonces, $x \in E_m$ para algún $m \in \mathbb{N}$,

$$\delta_x\left(\bigcup_{n=1}^{\infty} E_n\right) = \delta_x(\{x\}) = 1 = \delta_x(E_m) = \sum_{n=1}^{\infty} \delta_x(E_n).$$

Pues $\delta_x(E_n) = 0$ para toda $n \neq m$.
Caso 2: $x \notin \bigcup_{n=1}^{\infty} E_n$

$$\delta_x\left(\bigcup_{n=1}^{\infty} E_n\right) = \delta_x(\emptyset) = 0 = \sum_{n=1}^{\infty} \delta_x(E_n).$$

Pues $\delta_x(E_n) = 0$ para toda $n \in \mathbb{N}$. Así pues, en cualquier caso se tiene que $\delta_x(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \delta_x(E_n)$. Por lo tanto, δ_x es una medida. Más aún, pertenece al espacio $\mathcal{M}(X)$ pues,

$$\delta_x(X) = \delta_x(\{x\}) = 1.$$

Ahora bien, a partir de la medida δ_x definiremos una nueva medida. La nueva medida consistirá en una suma finita de medidas δ_x, como se muestra en la siguiente proposición.

Proposición 26. *Sea $m \in \mathbb{N}$. La función $\mu : S \to \mathbb{R}$ definida como sigue:*

$$\mu = \frac{1}{m} \sum_{k=0}^{m-1} \delta_{T^k(x)}.$$

donde δ_x es la medida definida como $\delta_x(\{x\}) = 1$ es una medida que pertenece a $\mathcal{M}(X)$.

Demostración. Veamos que $\mu(\emptyset) = 0$.

$$\mu(\emptyset) = \frac{1}{m} \sum_{k=0}^{m-1} \delta_{T^k(x)}(\emptyset) = \frac{1}{m} \sum_{k=0}^{m-1} 0 = 0.$$

Sea $E \in S$. Veamos que $\mu(E) \geq 0$. Sabemos que $\delta_{T^k(x)}(E) \geq 0$ para toda $k \in \mathbb{N}$, entonces,

$$\sum_{k=0}^{m-1} \delta_{T^k(x)}(E) \geq 0.$$

Así también,

$$\mu(E) = \frac{1}{m} \sum_{k=0}^{m-1} \delta_{T^k(x)}(E) \geq 0.$$

Sea $(E_n)_{n\in\mathbb{N}}$ una sucesión de elementos disjuntos de S. Veamos que $\mu(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \mu(E_n)$. Por definición se tiene lo siguiente,

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \frac{1}{m} \sum_{k=0}^{m-1} \delta_{T^k(x)}\left(\bigcup_{n=1}^{\infty} E_n\right).$$

Pero como $\delta_{T^k(x)}$ es una medida para cada $k \in \mathbb{N}$, entonces se tiene que $\delta_{T^k(x)}(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \delta_{T^k(x)}(En)$ para cada $k \in \mathbb{N}$. Entonces,

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} E_n\right) &= \frac{1}{m}\sum_{k=0}^{m-1}\sum_{n=1}^{\infty} \delta_{T^k(x)}(En) \\
&= \frac{1}{m}\sum_{n=1}^{\infty}\sum_{k=0}^{m-1} \delta_{T^k(x)}(En) \\
&= \sum_{n=1}^{\infty}\frac{1}{m}\sum_{k=0}^{m-1} \delta_{T^k(x)}(En) \\
&= \sum_{n=1}^{\infty} \mu(E_n).
\end{aligned}$$

Por lo tanto, concluimos que μ es una medida. Más aún, pertenece al espacio $\mathcal{M}(X)$, pues,

$$\begin{aligned}
\mu(X) &= \frac{1}{m}\sum_{k=0}^{m-1} \delta_{T^k(x)}(X) \\
&= \frac{1}{m}\sum_{k=0}^{m-1} 1 \\
&= \frac{m}{m} = 1.
\end{aligned}$$

□

En la siguiente proposición definimos una sucesión de medidas $(\mu_n)_{n\in\mathbb{N}}$ para un espacio medible (X, S) y una transformación T, cuyos puntos de acumulación μ son tales que T preserva la medida μ.

Proposición 27. *Sea (X, ρ) un espacio métrico compacto y $T : X \to X$ una transformación continua. Para cada $x \in X$, la sucesión de medidas en $\mathcal{M}(X)$*

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}$$

tiene al menos un punto de acumulación en $\mathcal{M}(X)$ y cualquiera de los puntos de acumulación son medidas en el espacio $\mathcal{M}(X)$, que son preservadas por la transformación T.

Demostración. Por hipótesis, X es un espacio métrico compacto, entonces por el Teorema 8 tenemos que $\mathcal{M}(X)$ es compacto. Así pues, para cada sucesión de medidas en $\mathcal{M}(X)$ existe una subsucesión convergente en $\mathcal{M}(X)$. Como $\delta_{T^k(x)} \in \mathcal{M}(X)$, entonces tenemos que cada $\mu_n \in \mathcal{M}(X)$. Así, $(\mu_n)_{n\in\mathbb{N}}$ tiene una subsucesión convergente, digamos $(\mu_{m_n})_{n\in\mathbb{N}}$ que converge a $\nu \in \mathcal{M}(X)$.

Consideremos la transformación $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ definida como,

$$\mu \mapsto \mu \circ T^{-1}$$
$$(T_*\mu)(A) = \mu(T^{-1}(A)).$$

En particular, $(T_*\delta_x)(A) = \delta_x(T^{-1}(A))$ pero,

$$\begin{aligned}\delta_x(T^{-1}(A)) = 1 &\Leftrightarrow x \in T^{-1}(A)\\ &\Leftrightarrow T(x) \in A\\ &\Leftrightarrow \delta_{T(x)}(A) = 1.\end{aligned}$$

Es decir, $(T_*\delta_x)(A) = \delta_{T(x)}(A)$. En general,

$$(T_*^k\delta_x)(A) = \delta_{T^k(x)}(A) \qquad \forall k \in \mathbb{N}. \tag{3.3}$$

Así pues, tenemos lo siguiente,

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1}\delta_{T^k(x)} = \frac{1}{n}\sum_{k=0}^{n-1}T_*^k\delta_x.$$

Consideremos la subsucesión convergente $(\mu_{m_n})_{n\in\mathbb{N}}$ que converge a ν. Entonces,

$$\begin{aligned}T_*\mu_{m_n} &= \frac{1}{m_n}\sum_{k=1}^{m_n}T_*^k\delta_x\\ &= \frac{1}{m_n}\sum_{k=1}^{m_n}\delta_{T^k(x)}\\ &= \mu_{m_n} + \frac{T_*^{m_n}\delta_x}{m_n} - \frac{\delta_x}{m_n}\\ &= \mu_{m_n} + \frac{\delta_{T^{m_n}(x)}}{m_n} - \frac{\delta_x}{m_n}.\end{aligned}$$

Haciendo $n \to \infty$ y por la continuidad de T_* (Pues T es continua por hipótesis) se tiene que $T_*(\nu) = \nu$. Por lo tanto, $\nu \in \mathcal{M}(X)$ y preserva la medida T. □

El siguiente lema es un resultado auxiliar que se utilizará para probar que la convergencia en el Teorema Ergódico de Birkhoff es uniforme cuando T es una transformación continua y únicamente ergódica en un espacio métrico compacto.

Lema 10. *Sea $n \in \mathbb{N}$. Sea (X, μ_n, S) un espacio de medida con μ_n definida como en la Proposición 27. Sea $\phi \in L^1(X, \mu_n)$, entonces se cumple lo siguiente:*

$$\int \phi \mathrm{d}\mu_n = \frac{1}{n}\sum_{k=0}^{n-1}\phi(T^k(x)).$$

Demostración. Sea $B \in S$ y consideremos $\phi = \chi_B$. Entonces,

$$\begin{aligned}\int \chi_B \mathrm{d}\mu_n &= \mu_n(B) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}(B).\end{aligned}$$

Pero $\delta_{T^k(x)}(B) = \chi_B(T^k(x))$, pues $\chi_B(T^k(x)) = 1$ si y sólo si $T^k(x) \in B$, pero esto ocurre si y sólo si $\delta_{T^k(x)}(B) = 1$. Entonces,

$$\int \chi_B \mathrm{d}\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \chi_B(T^k(x)).$$

Ahora bien, sea s una función simple. Entonces s tiene la forma $s = \sum_{i=1}^{N} \alpha_i \chi_{E_i}$ con $\alpha_i \in \mathbb{R}$, E_i en el álgebra de conjuntos $\forall i \in \{1, \cdots, N\}$ y $N \in \mathbb{N}$.

$$\begin{aligned}\int s \mathrm{d}\mu_n &= \int \sum_{i=1}^{N} \alpha_i \chi_{E_i} \mathrm{d}\mu_n \\ &= \sum_{i=1}^{N} \alpha_i \int \chi_{E_i} \mathrm{d}\mu_n \\ &= \sum_{i=1}^{N} \alpha_i \frac{1}{n}\sum_{k=0}^{n-1} \chi_{E_i}(T^k(x)) \\ &= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{i=1}^{N} \alpha_i \chi_{E_i}(T^k(x)) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} s(T^k(x)).\end{aligned}$$

Ahora, sea f una función medible. Por el Lema Básico de Aproximación, se tiene que existe una sucesión de funciones simples $(s_m)_{m\in\mathbb{N}}$ tales que $\lim\limits_{m\to\infty} s_m = f$. Por el Teorema de la Convergencia Monótona,

$$\begin{aligned}\int f \mathrm{d}\mu_n &= \int \lim_{m\to\infty} s_m \mathrm{d}\mu_n \\ &= \lim_{m\to\infty} \int s_m \mathrm{d}\mu_n \\ &= \lim_{m\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} s_m(T^k(x)) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \lim_{m\to\infty} s_m(T^k(x)) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} f(T^k(x)).\end{aligned}$$

Sea $\phi \in L^1(X, \mu_n)$, entonces existen ϕ^+, ϕ^- funciones medibles tales que $\phi = \phi^+ - \phi^-$. Por la linealidad de la integral, tenemos lo siguiente,

$$\begin{aligned}\int \phi \mathrm{d}\mu_n &= \int \phi^+ \mathrm{d}\mu_n - \int \phi^- \mathrm{d}\mu_n \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \phi^+(T^k(x)) - \frac{1}{n}\sum_{k=0}^{n-1} \phi^-(T^k(x)) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} (\phi^+(T^k(x)) - \phi^-(T^k(x))) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \phi(T^k(x)).\end{aligned}$$

□

A continuación probaremos el resultado principal de esta sección. Si T es continua y únicamente ergódica, la convergencia en el Teorema Ergódico de Birkhoff es uniforme.

Teorema 24. *Sea (X, ρ) un espacio métrico compacto y $T : X \to X$ una transformación continua únicamente ergódica. Si $\phi \in C(X)$, entonces la sucesión de funciones,*

$$\frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k$$

converge uniformemente cuando $n \to \infty$ a la constante $\int \phi \mathrm{d}\mu$, donde $\{\mu\} = \mathcal{M}(X)$.

Demostración. Sea $x \in X$. Consideremos la siguiente sucesión de funciones:

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k(x)}$$

Por la Proposición 26, sabemos que μ_n es una medida para cada $n \in \mathbb{N}$. Así también, por la Proposición 27, se tiene que $(\mu_n)_{n\in\mathbb{N}}$ tiene al menos un punto de acumulación en $\mathcal{M}(X)$ y que cada punto de acumulación es una medida que pertenece a $\mathcal{M}(X)$ y es preservada por la transformación T.

Por hipótesis, $T : X \to X$ es únicamente ergódica, entonces T preserva una única medida $\mu \in \mathcal{M}(X)$. Así pues, toda subsucesión convergente de $(\mu_n)_{n\in\mathbb{N}}$ converge a μ. Sea $\phi \in C(X)$. Por el Lema 10 se tiene que,

$$\int \phi \mathrm{d}\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \phi(T^k(x)).$$

Como $\mu_n \to \mu$. Por la Proposición 12, se tiene que,

$$\int \phi \mathrm{d}\mu_n \to \int \phi \mathrm{d}\mu.$$

Entonces,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1}\phi(T^k(x)) = \int \phi \mathrm{d}\mu.$$

Ahora veamos que la convergencia es uniforme. Por contradicción, supongamos que la convergencia no es uniforme, es decir, para $\epsilon > 0$ y para cada $n \in \mathbb{N}$, se tiene que existe $m_n \geq n$ tal que,

$$\left\| \frac{1}{m_n}\sum_{k=0}^{m_n-1}\phi\circ T^k - \int \phi \mathrm{d}\mu \right\|_\infty \geq \epsilon \qquad \forall n \in \mathbb{N}.$$

Por otra parte, consideremos la siguiente sucesión de medidas.

$$\nu_n = \frac{1}{m_n}\sum_{k=0}^{m_n-1}\delta_{T^k(x_n)}.$$

Como X, por hipótesis es compacto, por el Teorema 8, tenemos que $\mathcal{M}(X)$ es compacto. Así pues, existe una subsucesión $(\nu_{l_n})_{n\in\mathbb{N}}$ de $(\nu_n)_{n\in\mathbb{N}}$ convergente a una medida $\nu \in \mathcal{M}(X)$. Veamos que T preserva la medida ν. Sea $\psi \in C(X)$, entonces,

$$\begin{aligned}
\int (\psi\circ T)\mathrm{d}\nu &= \lim_{n\to\infty}\int(\psi\circ T)\mathrm{d}\nu_{l_n} \\
&= \lim_{n\to\infty}\frac{1}{m_{l_n}}\sum_{k=0}^{m_{l_n}-1}\psi(T^{k+1}(x_{l_n})) \\
&= \lim_{n\to\infty}\frac{1}{m_{l_n}}\left(\sum_{k=0}^{m_{l_n}-1}\psi(T^k(x_{l_n})) + \psi(T^{m_{l_n}}(x_{l_n})) - \psi(x_{l_n})\right) \\
&= \lim_{n\to\infty}\frac{1}{m_{l_n}}\left(\sum_{k=0}^{m_{l_n}-1}\psi(T^k(x_{l_n})\right) \\
&= \int \psi \mathrm{d}\nu.
\end{aligned}$$

Por lo tanto,

$$\int (\psi\circ T)\mathrm{d}\nu = \lim_{n\to\infty}\int(\psi\circ T)\mathrm{d}\nu_{l_n} = \int\psi\mathrm{d}\nu.$$

Por la Proposición 15, tenemos que T preserva la medida ν y como por hipótesis se tiene que T es únicamente ergódica, concluimos que $\nu = \mu$.

$$\epsilon \leq \left| \frac{1}{m_{l_n}}\left(\sum_{k=0}^{m_{l_n}-1}\phi(T^k(x_{l_n})\right) - \int\phi\mathrm{d}\mu \right| = \left|\int \phi\mathrm{d}\nu_{l_n} - \int\phi\mathrm{d}\mu\right| \to 0.$$

Lo cual es una contradicción. Por lo tanto tenemos que la convergencia es uniforme. $\square$

Ahora probaremos que en el caso en el que el espacio X es métrico compacto y T una transformación continua, entonces se cumple el recíproco del Teorema 24. En la siguiente sección, utilizaremos este resultado para probar un criterio que nos diga cuándo una transformación es únicamente ergódica. La prueba ocupa el Teorema de Representación de Riesz que enunciamos anteriormente como el Teorema 7.

Teorema 25. *Sea (X,ρ) un espacio métrico compacto y $T : X \to X$ una transformación continua. Si para cada $\phi \in C(X)$, tenemos que la sucesión de funciones:*

$$\frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k,$$

converge uniformemente a una constante cuando $n \to \infty$, entonces la transformación T es únicamente ergódica.

Demostración. Definimos una funcional $J : C(X) \to \mathbb{R}$ como sigue:

$$J(\phi) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k.$$

Observamos que la funcional J está bien definida porque por hipótesis tenemos que $\frac{1}{n}\sum_{k=0}^{n-1} \phi\circ T^k$ converge uniformemente a una constante cuando $n \to \infty$ para toda función $\phi \in C(X)$. Así también, observamos que J es una funcional lineal pues,

$$\begin{aligned} J(\phi+\psi) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (\phi+\psi) \circ T^k \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k + \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (\psi) \circ T^k \\ &= J(\phi) + J(\psi). \end{aligned}$$

También,

$$\begin{aligned} J(\lambda\phi) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (\lambda\phi) \circ T^k \\ &= \lambda \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k \\ &= \lambda J(\phi). \end{aligned}$$

Veamos que cumple las hipótesis del Teorema de Representación de Riesz, es decir, que es positiva, continua y $J(1) = 1$. Para ver que es positiva, supongamos que $\phi \geq 0$ y veamos que $J(\phi) \geq 0$.

$$\begin{aligned} \phi \geq 0 \Rightarrow (\phi \circ T^k) &\geq 0 \qquad \forall k \in \mathbb{N} \\ \Rightarrow \sum_{k=0}^{n-1} (\phi \circ T^k) &\geq 0 \qquad \forall k \in \mathbb{N} \\ \Rightarrow \frac{1}{n} \sum_{k=0}^{n-1} (\phi \circ T^k) &\geq 0 \qquad \forall k \in \mathbb{N}. \end{aligned}$$

Así pues. $J(\phi) = \frac{1}{n}\sum_{k=0}^{n-1}(\phi \circ T^k) \geq 0$. Sea $\epsilon \geq 0$. Para ver que J es continua, tomamos $\delta = \epsilon$ y $\phi, \psi \in C(X)$ tales que $\|\phi - \psi\|_\infty \leq \delta$ y veamos que $|J(\phi) - J(\psi)|) \leq \epsilon$. Por linealidad,

$$\begin{aligned} |J(\phi) - J(\psi)| &= |J(\phi - \psi)| \\ &= \left| \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\phi + \psi) \circ T^k \right| \\ &= \lim_{n\to\infty} \frac{1}{n} \left| \sum_{k=0}^{n-1} (\phi + \psi) \circ T^k \right| \\ &\leq \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |(\phi + \psi) \circ T^k| \\ &\leq \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \|(\phi + \psi) \circ T^k\|_\infty \\ &= \lim_{n\to\infty} \|(\phi + \psi) \circ T^k\|_\infty \\ &= \|(\phi + \psi) \circ T^k\|_\infty \\ &< \delta = \epsilon. \end{aligned}$$

Entonces, J es continua. Finalmente, veamos que $J(1) = 1$.

$$\begin{aligned} J(1) &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (1) \circ T^k \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1 \\ &= \lim_{n\to\infty} 1 \\ &= 1. \end{aligned}$$

Ahora bien, por el Teorema de Representación de Riesz, tenemos que existe una única medida $\mu \in \mathcal{M}(X)$ tal que,

$$J(\phi) = \int \phi \mathrm{d}\mu \qquad \forall \phi \in C(X). \tag{3.4}$$

Por otra parte, veamos que T preserva la medida μ. Observamos lo siguiente,

$$\begin{aligned} J(\phi \circ T) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^{k+1} \\ &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} (\phi \circ T^{k} + \phi \circ T^{n} - \phi) \\ &= J(\phi). \end{aligned}$$

Es decir, $J(\phi) = J(\phi \circ T)$ para toda función $\phi \in C(X)$. Por la Proposición 15 y por la observación (3.4) tenemos que T preserva la medida μ. Consideremos la siguiente sucesión de funciones $\phi_n : X \to \mathbb{R}$ definida como,

$$\phi_n = \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^{k}.$$

Como $\lim\limits_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^{k} = J(\phi)$ y $J(\phi) = \int \phi \mathrm{d}\mu$ se tiene,

$$\phi_n \to \int \phi \mathrm{d}\mu \qquad \text{si } n \to \infty.$$

Más aún, por hipótesis tenemos que la convergencia es uniforme. Finalmente veamos que μ es única. Supongamos que existe una medida ν que es preservada por la transformación T. Entonces se tiene lo siguiente,

$$\int \phi \mathrm{d}\nu \to \int\int \phi \mathrm{d}\mu \mathrm{d}\nu = \int \phi \mathrm{d}\mu \qquad \text{si } n \to \infty.$$

Como T preserva la medida ν, por la Proposición 15 se tiene,

$$\int (\phi \circ T)\mathrm{d}\nu = \int \phi \mathrm{d}\nu.$$

Entonces,

$$\begin{aligned} \int \phi_n \mathrm{d}\nu &= \int \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^{k}\mathrm{d}\nu \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \int \phi \circ T^{k}\mathrm{d}\nu \\ &= \frac{1}{n}\sum_{k=0}^{n-1} \int \phi \mathrm{d}\nu \\ &= \int \phi \mathrm{d}\nu. \end{aligned}$$

Por lo tanto, $\int \phi \mathrm{d}\nu = \int \phi \mathrm{d}\mu$ para toda función $\phi \in C(X)$, por lo que concluimos que $\nu = \mu$ y entonces T es únicamente ergódica. $\square$

3.3.2. Criterio de Ergodicidad Única

En esta sección veremos un criterio que nos permitirá saber si una transformación es únicamente ergódica. El siguiente teorema muestra que, si el espacio X es métrico y compacto, para probar que una transformación dada es únicamente ergódica, basta considerar las subfamilias F de funciones continuas que son densas en $C(X)$. Si las funciones en $\phi \in F$ satisfacen que la sucesión,

$$\frac{1}{n}\sum_{k=0}^{n-1}\phi \circ T^k.$$

converge uniformemente a una constante, entonces la transformación T es únicamente ergódica. Consideremos Φ una familia de funciones $\Phi \subseteq C(X)$. Denotamos por $L(\Phi)$ al conjunto de las combinaciones lineales de funciones en Φ.

Teorema 26. *Sea (X, ρ) un espacio métrico compacto y $T : X \to X$ una transformación continua. Sea $\Phi \subseteq C(X)$ una familia de funciones tales que $L(\Phi)$ es denso en $C(X)$. Si se tiene que para cada $\phi \in \Phi$ la sucesión,*

$$\frac{1}{n}\sum_{k=0}^{n-1}\phi \circ T^k$$

converge uniformemente a una constante cuando $n \to \infty$, entonces T es únicamente ergódica.

Demostración. Sea $\phi \in C(X)$. Por hipótesis $L(\Phi)$ es denso en $C(X)$, entonces existe una sucesión $(\psi_m)_{m\in\mathbb{N}}$ tal que $\psi_m \to \phi$ cuando $n \to \infty$, es decir,

$$\|\phi - \psi_m\|_\infty \to 0 \qquad \text{si } n \to \infty.$$

Definimos $\psi_{mn} : X \to \mathbb{R}$ como sigue,

$$\psi_{mn} = \frac{1}{n}\sum_{k=0}^{n-1}\psi_m \circ T^k.$$

Veamos que para cada $m \in \mathbb{N}$, existe $c_m \in \mathbb{R}$ tal que $\|\phi - \psi_m\|_\infty \to$ si $n \to \infty$. Sea $m \in \mathbb{N}$. Nos fijamos en $\psi_m \in L(\Phi)$. Como ψ_m es combinación lineal de elementos de Φ, entonces tenemos que existen,

$$\begin{aligned}&\eta_1, \eta_2, \cdots, \eta_p \in \Phi\\ &a_{m_1}, a_{m_2}, \cdots, a_{m_p} \in \mathbb{R}.\end{aligned}$$

Tales que,

$$\psi_m = a_{m_1}\eta_1 + a_{m_2}\eta_2 + \cdots + a_{m_p}\eta_p.$$

Por hipótesis, tenemos que para cada $f \in \Phi$, la sucesión

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow{u} c = constante.$$

Así pues, en particular se tiene que,

$$\frac{1}{n}\sum_{k=0}^{n-1}\eta_j \circ T^k \xrightarrow{u} d_j \in \mathbb{R} \qquad \forall j \in \{1, 2, \cdots p\}.$$

Definimos,

$$c_m = a_{m_1}d_1 + a_{m_2}d_2 + \cdots + a_{m_p}d_p.$$

Entonces, $\|\psi_{mn} - c_m\|_\infty$ es igual a,

$$\begin{aligned}
&= \left\|\frac{1}{n}\sum_{k=0}^{n-1}\psi_m \circ T^k - c_m\right\|_\infty \\
&= \left\|\frac{1}{n}\sum_{k=0}^{n-1}(a_{m_1}\eta_1 + \cdots + a_{m_p}\eta_p) \circ T^k - c_m\right\|_\infty \\
&= \left\|a_{m_1}\frac{1}{n}\sum_{k=0}^{n-1}\eta_1 \circ T^k + \cdots + a_{m_p}\frac{1}{n}\sum_{k=0}^{n-1}\eta_p \circ T^k - c_m\right\|_\infty \\
&= \left\|a_{m_1}\frac{1}{n}\sum_{k=0}^{n-1}\eta_1 \circ T^k + \cdots + a_{m_p}\frac{1}{n}\sum_{k=0}^{n-1}\eta_p \circ T^k - a_{m_1}d_1 - \cdots - a_{m_p}d_p\right\|_\infty \\
&= \left\|a_{m_1}\left(\frac{1}{n}\sum_{k=0}^{n-1}\eta_1 \circ T^k - d_1\right) + \cdots + a_{m_p}\left(\frac{1}{n}\sum_{k=0}^{n-1}\eta_p \circ T^k - a_{m_1}d_1 - d_p\right)\right\|_\infty \\
&\leq |a_{m_1}|\left\|\frac{1}{n}\sum_{k=0}^{n-1}\eta_1 \circ T^k - d_1\right\|_\infty + \cdots + |a_{m_p}|\left\|\frac{1}{n}\sum_{k=0}^{n-1}\eta_p \circ T^k - a_{m_1}d_1 - d_p\right\|_\infty \\
&\to 0 \qquad \text{si } n \to \infty.
\end{aligned}$$

Luego pues,

$$\|\psi_{mn} - c_m\|_\infty \to 0 \qquad \text{si } n \to \infty. \tag{3.5}$$

Ahora bien, sean $m, l \in \mathbb{N}$.

$$\begin{aligned}
\|\psi_{mn} - \psi_{ln}\|_\infty &\leq \frac{1}{n}\sum_{k=0}^{n-1}\|\psi_m \circ T^k - \psi_l \circ T^k\|_\infty \\
&\leq \frac{1}{n}\sum_{k=0}^{n-1}\|\psi_m - \psi_l\|_\infty \\
&= \|\psi_m - \psi_l\|_\infty.
\end{aligned}$$

Más aún, tomando en cuenta la afirmación anterior y por la desigualdad del triángulo,

$$\begin{aligned}
|c_m - c_l| &\leq \|c_m - \psi_{mn}\|_\infty + \|\psi_{mn} - \psi_{ln}\|_\infty + \|\psi_{ln} - c_l\|_\infty \\
&\leq \|c_m - \psi_{mn}\|_\infty + \|\psi_m - \psi_l\|_\infty + \|\psi_{ln} - c_l\|_\infty.
\end{aligned}$$

Haciendo $n \to \infty$, se tiene que,

$$|c_m - c_l| \leq \|\psi_m - \psi_l\|_\infty. \tag{3.6}$$

Como $(\psi_m)_{m\in\mathbb{N}}$ es convergente, entonces tenemos que es de Cauchy y por la desigualdad anterior (3.6), se sigue que $(c_m)_{m\in\mathbb{N}}$ es de Cauchy. Como $(\mathbb{R}, |\cdot|)$ es completo, tenemos que $(c_m)_{m\in\mathbb{N}}$ es convergente. Supongamos que $(c_m)_{m\in\mathbb{N}}$ converge a $c \in \mathbb{R}$. Ahora, definimos la siguiente sucesión de funciones,

$$\phi_n = \frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k \qquad n \in \mathbb{N}.$$

Veamos que $\phi_n \xrightarrow{u} c$ cuando $n \to \infty$. Observamos que para alguna $m \in \mathbb{N}$ se tiene,

$$\|\phi_n - c\|_\infty \leq \|\phi_n - \psi_{mn}\|_\infty + \|\psi_{mn} - c_m\|_\infty + |c_m - c|.$$

Afirmación:$\|\phi_n - \psi_{mn}\|_\infty \leq \|\phi_n - \psi_m\|_\infty$.

$$\begin{aligned}
\|\phi_n - \psi_{mn}\|_\infty &= \left\|\frac{1}{n}\sum_{k=0}^{n-1}\phi \circ T^k - \frac{1}{n}\sum_{k=0}^{n-1}\psi_m \circ T^k\right\|_\infty \\
&= \left\|\frac{1}{n}\sum_{k=0}^{n-1}(\phi - \psi_m)\circ T^k\right\|_\infty \\
&\leq \frac{1}{n}\sum_{k=0}^{n-1}\|(\phi - \psi_m)\circ T^k\|_\infty \\
&\leq \frac{1}{n}\sum_{k=0}^{n-1}\|\phi - \psi_m\|_\infty \\
&= \|\phi - \psi_m\|_\infty.
\end{aligned}$$

Así pues,

$$\|\phi_n - c\|_\infty \leq \|\phi - \psi_m\|_\infty + \|\psi_{mn} - c_m\|_\infty + |c_m - c|.$$

Como $\psi_m \to \phi$ y $c_m \to c$ cuando $m \to \infty$, para $\epsilon > 0$ y m suficientemente grande se tiene que,

$$\|\phi - \psi_m\|_\infty < \epsilon \text{ y } \quad |c_n - c| < \epsilon.$$

Sabemos por (3.5) que $\|\psi_{mn} - c_m\|_\infty \to 0$ cuando $n \to \infty$. Entonces,

$$\limsup_{n\to\infty}\|\phi_n - c\|_\infty \leq 2\epsilon.$$

Como ϵ es arbitrario, concluimos,

$$\phi_n \xrightarrow{u} c \qquad \text{si } n \to \infty.$$

Por el criterio de ergodicidad única que se probó en el Teorema 25, T es únicamente ergódica. $\square$

A continuación veremos como ejemplo de una transformación únicamente ergódica a las rotaciones irracionales, es decir, las rotaciones donde $w = e^{2\pi i\alpha}$ con α irracional. Para ello, probaremos primero que las traslaciones irracionales son únicamente ergódicas. Finalmente, para terminar el capítulo, veremos una aplicación a la Teoría de los Números.

3.4. Rotaciones en el Círculo

En la sección 3.2.2 probamos que las rotaciones irracionales R_w son ergódicas, mientras que las rotaciones racionales no lo son. En esta sección se mostrará no sólo que las rotaciones irracionales son ergódicas, sino que son únicamente ergódicas. Al ser únicamente ergódicas tenemos que la medida de Lebesgue λ en $\mathbb{S}^1$ es la única medida que preserva a las rotaciones irracionales.

Para probar que las rotaciones irracionales son únicamente ergódicas ocuparemos el Teorema 26. Así pues, debemos encontrar una familia de funciones tal que sus combinaciones lineales sean densas sobre el espacio de las funciones continuas sobre $\mathbb{S}^1$. Lo primero que observamos es que en el espacio $[0,1]$, sabemos que los polinómios trigonométricos son densos sobre el espacio $C([0,1])$ de las funciones continuas sobre el $[0,1]$. Es decir, si se toma la familia,

$$\Phi = \{1\} \cup \{\cos(2\pi kx) : k \in \mathbb{N}\} \cup \{\sin(2\pi kx) : k \in \mathbb{N}\}.$$

Sabemos que $L(\Phi)$ es densa en $C([0,1])$. Consideremos a las rotaciones R_w y a las funciones $\chi_k(z) = z^k$ para cada $k \in \mathbb{Z}$. Veremos que las combinaciones lineales de la familia,

$$\Phi = \{\chi_k : k \in \mathbb{N}\}$$

es densa sobre el espacio $C(\mathbb{S}^1)$. Tomamos $\chi_0 = 1$ y observamos que

$$\begin{aligned}\chi_k(e^{2\pi ix}) + \chi_{-k}(e^{2\pi ix}) &= e^{2\pi ixk} + e^{-2\pi ixk}\\ &= \cos 2\pi xk + i\sin 2\pi xk + \cos 2\pi xk - i\sin 2\pi xk\\ &= 2\cos 2\pi xk.\end{aligned}$$

También,

$$\begin{aligned}\chi_k(e^{2\pi ix}) - \chi_{-k}(e^{2\pi ix}) &= e^{2\pi ixk} - e^{-2\pi ixk}\\ &= \cos 2\pi xk + i\sin 2\pi xk - \cos 2\pi xk + i\sin 2\pi xk\\ &= 2i\sin 2\pi xk.\end{aligned}$$

Por la observación anterior, concluimos que las combinaciones lineales de la familia de funciones $\Phi = \{\chi_k : k \in \mathbb{N}\}$ son densas en $C(S^1)$. Consideremos la rotación irracional R_w. Supongamos primero que $k \neq 0$. Esto implica que $w^k \neq 1$ para cada k entero distinto del 0, la sucesión,

$$\frac{1}{n}\sum_{j=0}^{n-1} \chi_k \circ R_w^j$$

converge uniformemente a una constante si $n \to \infty$., En efecto,

$$\begin{aligned}
\left|\frac{1}{n}\sum_{j=0}^{n-1}\chi_k(R_w^j(z))\right| &= \left|\frac{1}{n}\sum_{j=0}^{n-1}\chi_k(w^j z)\right| \\
&= \left|\frac{1}{n}\sum_{j=0}^{n-1} w^{jk} z^k\right| \\
&= \left|\frac{1}{n}\sum_{j=0}^{n-1} w^{jk}\chi_k(z)\right| \\
&= \frac{1}{n}\left|\sum_{j=0}^{n-1} w^{jk}\right| |\chi_k(z)|\,.
\end{aligned}$$

Como $|\chi_k(z)| = 1$, entonces,

$$\begin{aligned}
\left|\frac{1}{n}\sum_{j=0}^{n-1}\chi_k(R_w^j(z))\right| &= \frac{1}{n}\frac{|1-w^{kn}|}{|1-w^k|} \\
&\leq \frac{1}{n}\frac{2}{|1-w^k|} \to 0 \qquad \text{si } n \to \infty.
\end{aligned}$$

Supongamos ahora que $k = 0$, entonces,

$$\frac{1}{n}\sum_{j=0}^{n-1}\chi_0(R_w^j(z)) = \frac{1}{n}\sum_{j=0}^{n-1} 1 = 1.$$

Para cada $\chi_k \in \Phi$, tenemos que $\frac{1}{n}\sum_{j=0}^{n-1}\chi_k(R_w^j(z))$ converge uniformemente a una constante. Entonces, por el Teorema 26, se concluye que las rotaciones irracionales son únicamente ergódicas y así, la única medida que preservan las rotaciones irracionales es la medida de Lebesgue sobre $\mathbb{S}^1$.

Para ver que las traslaciones irracionales son únicamente ergódicas, utilizaremos el isomorfísmo $h : \Pi \to \mathbb{S}^1$ definido como $h(\theta) = e^{2\pi i\theta}$. Sea T_α una traslación irracional, como,

$$R_w \circ h = h \circ T_\alpha.$$

Entonces,

$$T_\alpha = h^{-1} \circ R_w \circ h.$$

Para $f \in C(\Pi)$ se tiene la sucesión,

$$\frac{1}{n}\sum_{j=0}^{n-1} f(T_\alpha(x)^j) = \frac{1}{n}\sum_{j=0}^{n-1} f(h^{-1} \circ R_w^j \circ h(x)).$$

Como h es continua y biyectiva, entonces $f \circ h^{-1} \in C(\mathbb{S}^1)$. Como sabemos que las rotaciones irracionales son únicamente ergódicas, Por el Teorema 24, tenemos que la sucesión,

$$\frac{1}{n}\sum_{j=0}^{n-1} f \circ h^{-1}(R_w^j \circ h(x))$$

converge uniformemente a una constante. Finalmente, por el Teorema 25 se concluye que las traslaciones irracionales con únicamente ergódicas.

3.5. Primer Dígito de las Potencias de 2

Ahora veamos una aplicación de la ergodicidad única a la Teoría de Números. Consideremos la sucesión de dígitos:

$$(l_n)_{n\in\mathbb{N}} = \{1, 2, 4, 8, 1, 3, 6, \cdots\},$$

que se obtienen al tomar el primer dígito de la sucesión:

$$\{2^n : n \in \mathbb{N}\} = \{1, 2, 4, 8, 16, 32, 64, \cdots\}.$$

Queremos encontrar la frecuencia relativa con la que cada dígito $k \in \{1, 2, \cdots, 9\}$ aparece en la sucesión $(l_n)_{n\in\mathbb{N}}$. Consideremos la Traslación del intervalo,

$$T(x) = x + \log_{10} 2 - \lfloor x + \log_{10} 2\rfloor.$$

Como $\log_{10} 2$ es irracional, entonces T es una trasformación únicamente ergódica. Observamos que el primer dígito de 2^n es k si y sólo si,

$$k \cdot 10^r \leq 2^n < (k+1)\cdot 10^r \qquad \text{para alguna } r \geq 0.$$

Así pues,

$$r + \log_{10} k \leq n \log_{10} 2 < r + \log_{10}(k+1).$$

Entonces $r = \lfloor n \log_{10} 2\rfloor$,

$$\log_{10} k \leq n\log_{10} 2 - \lfloor n\log_{10} 2\rfloor < \log_{10}(k+1)$$

$$\log_{10} k \leq T^n(0) < \log_{10}(k+1).$$

Por lo tanto, $k = l_n$ si y sólo si $T^n(0) \in [\log_{10} k, \log_{10}(k+1))$. Por la ergodicidad única de T, tenemos que para cada $k \in \{1, 2, \cdots, 9\}$,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\chi_{[\log_{10} k, \log_{10}(k+1))}(T^n(0)) &= m([\log_{10} k, \log_{10}(k+1))) \\ &= \log_{10}\left(\frac{k+1}{k}\right).\end{aligned}$$

Capítulo 4

Entropía

En este capítulo introduciremos la noción de *entropía* en tres niveles

- entropía de una partición medible.
- entropía con respecto a una partición medible y una transformación medible.
- entropía con respecto a una transformación medible.

Posteriormente probaremos el *Teorema de Kolmogorov-Sinaí*, el cuál nos facilita el cálculo de la entropía y por último probaremos el *Teorema de Shannon-McMillan-Breiman*.

4.1. Entropía de una Partición Medible

La noción de *entropía* fue propuesta por primera vez por C. E. Shannon en su artículo "La Teoría Matemática de la Comunicación"[1] al proponer un modelo para la transmisión de información. Sin embargo, sus propiedades y su carácter teórico han cobrado poco a poco relevancia en el estudio de los sistemas dinámicos y en otras áreas de las matemáticas.

La idea original consistía en partir de un *sistema completo de eventos* $A_1, A_2, \cdots, A_n$, es decir, un conjunto de eventos tales que uno y sólo uno de ellos ocurre en cada ensayo. Al considerarlos junto con sus probabilidades, se obtiene lo que llamaremos un *esquema finito*. Cada esquema finito describe un *estado de incertidumbre*. Por ejemplo, si consideramos los siguientes esquemas:

$$\begin{pmatrix} A_1 & A_2 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \qquad \begin{pmatrix} A_1 & A_2 \\ \frac{99}{100} & \frac{1}{100} \end{pmatrix}$$

Observamos que el primer esquema representa una mayor incertidumbre que el primero, pues al realizar un ensayo bajo el primer esquema, el evento A_1 tiene

[1] Shannon C. A. 1948, "A Mathematical Theory of Communication", *The Bell System Technical Journal*, Vol. 27, pp. 379-423.

la misma probabilidad de ocurrir que el evento A_2. Sin embargo, en el segundo esquema, se tiene que es casi seguro que ocurra el evento A_1. Para medir la "incertidumbre" en los esquemas finitos se utiliza la noción de entropía mediante una fórmula que recuerda el concepto de entropía tal y como se presenta en la termodinámica. Si $p_1, p_2, \cdots, p_n$ son las probabilidades respectivas de los eventos $A_1, A_2, \cdots, A_n$, entonces la entropía del esquema finito es la siguiente,

$$H(p_1, p_2, \cdots, p_n) = -\sum_{k=1}^{n} p_k \log p_k,$$

donde $p_k \log p_k = 0$ si $p_k = 0$. La función $H(p_1, p_2, \cdots, p_n)$ resulta adecuada porque preserva las propiedades razonables para medir la incertidumbre de un esquema finito. Por ejemplo, $H(p_1, p_2, \cdots, p_n) = 0$ en caso de que alguno de los números $p_1, p_2, \cdots, p_n$ sea uno y los demás cero, es decir, el esquema carece de incertidumbre dado que alguno de los eventos ocurrirá con certeza. En la sección 4.1.2 probaremos a detalle las propiedades mas importantes de la entropía.

4.1.1. Particiones Medibles

En esta sección abstraemos la noción de esquema finito y la formalizamos sobre lo que llamaremos "particiones medibles". El concepto se trabaja en general sobre cualquier espacio medible de medida finita. El concepto de partición medible se construye a partir del concepto de *pre-partición medible* mismo que definimos de manera formal a continuación.

Definición 23. *Sea (X, S, μ) un espacio medible tal que $\mu(X) = 1$. Sea $\xi \subseteq S$ una familia finita de subconjuntos de S. Decimos que ξ es una **pre-partición medible** de X (con respecto a μ) si,*

- $\mu(\bigcup_{C\in\xi} C) = 1$.
- $\mu(C \cap D) = 0$ *para* $C, D \in \xi$, $C \neq D$.

Ejemplo 12. Consideremos el espacio medible $([0,1], S, m)$, dónde S es la σ-álgebra de Borel y m es la medida de Lebesgue. Una pre-partición medible de $[0,1]$ sería la siguiente:

$$\xi = \left(0, \tfrac{1}{3}\right), \left(\tfrac{1}{3}, 1\right)$$

En lo que sigue, definiremos una relación $\sim$ entre las pre-particiones medibles de un espacio de medida X. Posteriormente probaremos que la relación definida es una relación de equivalencia.

Definición 24. *Sea (X, S, μ) un espacio medible tal que $\mu(X) = 1$. Sean ξ, η dos pre-particiones de X con respecto a μ. Decimos que $\xi \sim \eta$ si,*

- $\forall C \in \xi$, $\exists D \in \eta$ *tal que* $\mu(C \setminus D) = 0$.
- $\forall D \in \eta$, $\exists C \in \xi$ *tal que* $\mu(D \setminus C) = 0$.

La definición anterior afirma que dos pre-particiones están relacionadas si y sólo si sus elementos difieren a lo más en un conjunto de medida cero. Veamos que la relación $\sim$ es una relación de equivalencia.

Proposición 28. *La relación $\sim$ es una relación de equivalencia.*

Demostración. Para ver que la relación $\sim$ es una relación de equivalencia, debemos verificar que la relación es reflexiva, simétrica y transitiva. Veamos primero que $\sim$ es una relación reflexiva, es decir, veamos que $\xi \sim \xi$. La reflexividad se sigue directamente de la definición al observar que para cada $C \in \xi$ se tiene que,

$$\mu(C \setminus C) = \mu(\emptyset) = 0.$$

Veamos ahora que la relación $\sim$ es simétrica, es decir, si $\xi \sim \eta$, entonces $\eta \sim \xi$. Supongamos que $\xi \sim \eta$, entonces por definición tenemos,

- $\forall C \in \xi$, $\exists D \in \eta$ tal que $\mu(C \setminus D) = 0$.
- $\forall D \in \eta$, $\exists C \in \xi$ tal que $\mu(D \setminus C) = 0$.

Por lo que también $\eta \sim \xi$. Así, $\sim$ es simétrica. Finalmente, debemos verificar que la relación $\sim$ es transitiva, es decir, si $\xi \sim \eta$ y $\eta \sim \theta$, entonces $\xi \sim \theta$. Sea $C \in \xi$, veamos que existe $H \in \theta$ tal que $\mu(C \setminus H) = 0$. También, como $\xi \sim \eta$, entonces existe $D \in \eta$ tal que $\mu(C \setminus D) = 0$. Como también estamos suponiendo que $\eta \sim \theta$, entonces $\exists H \in \theta$ tal que $\mu(D \setminus H) = 0$. Veamos que $\mu(C \setminus H) = 0$.

$$\begin{aligned} C \setminus H &= (C \cap H^c) \cap X \\ &= (C \cap H^c) \cap (D \cup D^c) \\ &= (C \cap H^c \cap D) \cup (C \cap H^c \cap D^c)) \\ &= ((D \setminus H) \cap C) \cup ((C \setminus D) \cap H^c). \end{aligned}$$

Pero $(D \setminus H) \cap C \subseteq D \setminus H)$ y $(C \setminus D) \cap H^c \subseteq C \setminus D$.

$$C \setminus H \subseteq (D \setminus H) \cup (C \setminus D).$$

Esto implica que,

$$\begin{aligned} 0 \leq \mu(C \setminus H) &\leq \mu((D \setminus H) \cup (C \setminus D)) \\ &\leq \mu(D \setminus H) + \mu(C \setminus D) \\ &= 0 + 0. \end{aligned}$$

Por lo tanto $\mu(C \setminus H) = 0$. Análogamente se tiene que si $H \in \theta$, entonces existe $C \in \xi$ tal que $\mu(H \setminus C) = 0$. Entonces, $\sim$ es transitiva. Por lo tanto, $\sim$ es una relación de equivalencia. □

Ejemplo 13. Consideramos el espacio medible $([0, 1], S, m)$, dónde S es la σ-álgebra de Borel y m es la medida de Lebesgue. Las siguientes pre-particiones medibles son equivalentes.

- $\left(0, \frac{1}{3}\right), \left(\frac{1}{3}, 1\right)$
- $\left[0, \frac{1}{3}\right), \left[\frac{1}{3}, 1\right)$
- $\left[0, \frac{1}{3}\right], \left[\frac{1}{3}, 1\right]$
- $\left[0, \frac{1}{3}\right], \left(\frac{1}{3}, 1\right)$.

A cada clase de equivalencia inducida por la relación $\sim$ se le llama ***partición medible*** de X. Ahora bien, puesto que los elementos de cualquier pre-partición medible que representan una clase de equivalencia dada están bien definidos, salvo conjuntos de medida cero, en lo que sigue no se hará distinción entre dos pre-particiones medibles que representan la misma partición medible.

4.1.2. Entropía de Shannon

En la introducción al capítulo presentamos la definición de entropía que propuso C. E. Shannon en términos de la probabilidad de los eventos en un sistema completo. En esta sección generalizaremos la noción de entropía para medidas y particiones medibles de un espacio de medida finita.

Definición 25. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. La **entropía de Shannon** o **entropía de una partición medible** ξ de X respecto a la medida μ se define como,*

$$H_\mu(\xi) = -\sum_{C\in\xi} \mu(C)\log(\mu(C)). \tag{4.1}$$

El número $H_\mu(\xi)$ se puede interpretar como la *cantidad de información* que se obtiene de una partición medible ξ de un espacio X. Mientras más pequeña sea la partición del espacio X, mayor será la cantidad de información que se obtenga.

La primera observación que haremos es que si para algún elemento de la partición $C \in \xi$ se tiene que $\mu(C) = 1$, entonces $H_\mu(\xi) = 0$, lo cual se puede interpretar como que no se obtiene nada de información de la partición medible.

Ejemplo 14. Consideremos el espacio medible $([0,1], S, m)$, donde m es la medida de Lebesgue y S la σ-álgebra de Borel. Así pues, para $k \in \mathbb{N}$ definimos una partición regular y calculamos su entropía.

$$\xi_j = \left(\frac{j}{k}, \frac{j+1}{k}\right) \qquad \textit{para cada } j = 0, 1, \cdots, k-1.$$

De manera que el siguiente conjunto es una partición medible de $[0,1]$.

$$\xi = \{\xi_j : j = 0, 1, \cdots, k-1\}.$$

Calculamos la entropía de la partición ξ del espacio $[0,1]$.

$$\begin{aligned} H_m(\xi) &= -\sum_{j=0}^{k-1} \frac{1}{k}\log(\frac{1}{k}) \\ &= -\frac{1}{k}\log(\frac{1}{k})k \\ &= -\log(\frac{1}{k}) \\ &= \log(k). \end{aligned}$$

El ejemplo anterior no sólo muestra el valor de la entropía de una *partición regular* del espacio medible, sino que también es un ejemplo de cuándo la entropía alcanza su valor máximo. La siguiente proposición nos da una cota al valor de la entropía de una partición medible. Con una partición regular se puede afirmar que se tiene la mayor cantidad de información. Una manera un poco más clara de ver esto es considerando un número de eventos con igual probabilidad de realizarse, en este caso, se podría decir que se tiene la mayor incertidumbre con respecto al evento que ocurrirá después del ensayo. Para probar la siguiente proposición, definimos la función $\psi : [0,1] \to \mathbb{R}$ como sigue,

$$\psi(x) = \begin{cases} x\log x, & \text{para } 0 < x \leq 1 \\ 0, & \text{para } x = 0. \end{cases}$$

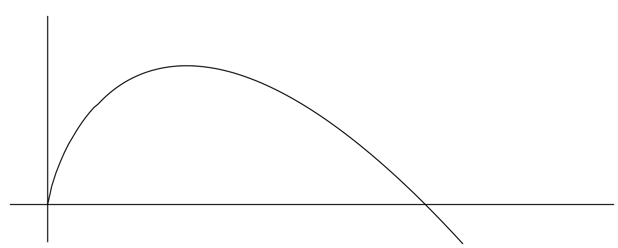


Figura 4.1: La función $f(x) = x\log x$ es una función convexa.

Notamos que podemos escribir a la entropía en términos de la función ψ de la siguiente manera,

$$H_\mu(\xi) = -\sum_{C\in\xi} \psi(\mu(C)).$$

Como $\psi''(x) = \frac{1}{x} > 0$, se tiene que ψ es una función estrictamente convexa. Entonces podemos aplicar la *desigualdad de Jensen*. Al aplicar la desigualdad de Jensen obtenemos lo siguiente,

$$\psi\left(\sum_{i=1}^{p} a_i x_i\right) \leq \sum_{i=1}^{p} a_i\psi(x_i),$$

donde $x_1, x_2, ..., x_p, a_1, a_2, ..., a_p \in [0,1]$ y $\sum_{i=1}^{p} a_i = 1$. La igualdad se cumple en el caso de que $x_i = a_i$ para cada i.

Proposición 29. *Sea (X, μ, S) un espacio de medida y sea ξ una partición medible, entonces*

$$H_\mu(\xi) \leq \log(|\xi|).$$

Demostración. Por definición se tiene,

$$H_\mu(\xi) = -\sum_{C\in\xi} \psi(\mu(C)) = -k \sum_{C\in\xi} \frac{1}{k}\psi(\mu(C)) \qquad \text{donde } k = |\xi|.$$

Como $\sum_{c\in\xi} \frac{1}{k} = |\xi| \frac{1}{k} = k\frac{1}{k} = 1$, por la desigualdad de Jensen tenemos,

$$\psi\left(\sum_{c\in\xi} \frac{1}{k}\mu(C)\right) \leq \sum_{C\in\xi} \frac{1}{k}\psi(\mu(C)).$$

Entonces,

$$-\sum_{C\in\xi} \frac{1}{k}\psi(\mu(C)) \leq -\psi\left(\sum_{c\in\xi} \frac{1}{k}\mu(C)\right).$$

Luego pues,

$$\begin{aligned} H_\mu(\xi) &\leq -k\psi\left(\sum_{C\in\xi} \frac{1}{k}\mu(C)\right) \\ &= -k\psi\left(\frac{\mu(\bigcup_{C\in\xi} C)}{k}\right) \\ &= -k\psi\left(\frac{1}{k}\right) \\ &= -\frac{k}{k}\log\frac{1}{k} \\ &= -\log\frac{1}{k} \\ &= \log k \\ &= \log(|\xi|). \end{aligned}$$

Por lo tanto,

$$H_\mu(\xi) \leq \log(|\xi|).$$

□

A continuación vamos a introducir la noción de *refinamiento de una partición medible*. Los refinamientos son particiones medibles que están contenidas en una partición medible mayor. De manera intuitiva, podríamos esperar que al obtener particiones mas finas cada vez, se obtenga cada vez más información del sistema.

Definición 26. *Sean ξ, η dos particiones medibles de X. Decimos que η es un* ***refinamiento*** *de ξ, o bien $\xi \leq \eta$ si $\forall D \in \eta$, $\exists C \in \xi$ tal que $\mu(D \setminus C) = 0$.*

La siguiente proposición muestra que la definición de refinamiento que se acaba de dar es adecuada, en el sentido de que recupera lo que intuitivamente esperaríamos que fuera un refinamiento si η es un refinamiento de ξ, entonces cada elemento de ξ se puede ver como la unión de elementos de η salvo conjuntos de medida cero.

Proposición 30. *Sea (X, S, μ) un espacio de medida y sean ξ, η dos particiones medibles de X. Si $\xi \leq \eta$, entonces,*

$$\mu(C) = \mu\left(\bigcup_{D \subseteq C} D\right) \qquad \forall C \in \xi.$$

Demostración. Por contradicción, supongamos que,

$$C \nsubseteq \bigcup_{D \subseteq C} D \qquad \textit{c.d. } (\text{rel } \mu).$$

Esto implica que existe $B \subseteq X$ conjunto medible con $\mu(B) > 0$ tal que,

$$\mu(C) = \mu\left(\bigcup_{D \subseteq C} D\right) + \mu(B).$$

Observamos que $B \subseteq C$ c.d. (relμ). Como η es una partición medible de X y por hipótesis es un refinamiento de ξ, para cada $D \in \eta$, existe $C'_D \in \xi$ tal que $D \subseteq C'_D$ c.d. rel(μ). Entonces,

$$B \subseteq \bigcup_{D \cap B \neq \emptyset} D \subseteq \bigcup_{D \cap B \neq \emptyset} C'_D \qquad \textit{c.d. } (\text{rel } \mu).$$

Esto implica que,

$$\begin{aligned} B \subseteq C \cap \bigcup_{D \cap B \neq \emptyset} C'_D \qquad \textit{c.d. } (\text{rel } \mu). \\ = \bigcup_{D \cap B \neq \emptyset} (C \cap C'_D) \qquad \textit{c.d. } (\text{rel } \mu). \end{aligned}$$

Finalmente, tomando medidas y del hecho de que ξ es una partición medible, se obtiene,

$$0 < \mu(B) \leq \sum_{D \cap B \neq \emptyset} \mu(C \cap C'_D) = 0.$$

Por lo tanto, $\mu(B) = 0$, lo cual es una contradicción. Luego pues, concluimos,

$$C \subseteq \bigcup_{D \subseteq C} D \qquad \textit{c.d. } (\text{rel } \mu).$$

Es decir,

$$\mu(C) = \mu\left(\bigcup_{D\subseteq C} D\right) \qquad \forall C \in \xi.$$

□

Ejemplo 15. Consideremos el espacio medible $([0,1], S, m)$ y sean $k, l \in \mathbb{N}$ tales que $k|l$. Definimos los conjuntos,

$$\xi_k = \left\{\left(\frac{j}{k}, \frac{j+1}{k}\right) : j = 0, 1, \cdots, k-1\right\}$$
$$\eta_l = \left\{\left(\frac{i}{l}, \frac{i+1}{l}\right) : i = 0, 1, \cdots, l-1\right\}.$$

Definimos también las particiones medibles inducidas por los conjuntos de arriba,

$$\xi = \{\xi_j : j = 0, 1, \cdots, k-1\}$$
$$\eta = \{\eta_i : i = 0, 1, \cdots, l-1\}.$$

Veamos que η es un refinamiento de ξ. Como supusimos que $k|l$, entonces tenemos que existe un entero q tal que $kq = l$.

$$\left(0, \frac{1}{k}\right) \Big\backslash \left(0, \frac{1}{l}\right) \cup \left(\frac{1}{l}, \frac{2}{l}\right) \cup \cdots \cup \left(\frac{l-1}{l}, \frac{l}{k}\right) = \emptyset.$$

Por el ejemplo anterior, sabemos el valor de la entropía.

$$H_m(\xi) = \log k \leq \log l = H_m(\eta).$$

El ejemplo anterior muestra a una partición y a su respectivo refinamiento. Más aun, al calcular la entropía de ambas particiones medibles se encontró que la entropía del refinamiento era mayor que el de la partición que refinaba. El siguiente resultado muestra que en general se cumple que la entropía de todo refinamiento es mayor que la entropía de la partición a la cual refina.

Proposición 31. *Sea (X, μ, S) un espacio de medida. Sean ξ y η dos particiones medibles de X. Si $\xi \leq \eta$, entonces $H_\mu(\xi) \leq H_\mu(\eta)$.*

Demostración. Por hipótesis, η es un refinamiento de ξ. Por la Proposición 30 tenemos que,

$$\mu(C) = \mu\left(\bigcup_{D\subseteq C} D\right) = \sum_{D\subseteq C} \mu(D) \qquad \forall C \in \xi.$$

Esto implica,

$$\begin{aligned}\psi(\mu(C)) &= \psi\left(\sum_{D\subseteq C}\mu(D)\right)\\ &= \sum_{D\subseteq C}\mu(D)\log\left(\sum_{D\subseteq C}\mu(D)\right)\\ &\geq \sum_{D\subseteq C}\mu(D)\log(\mu(D))\\ &= \sum_{D\subseteq C}\psi(\mu(D)).\end{aligned}$$

Es decir,

$$\begin{aligned}\sum_{D\subseteq C}\psi(\mu(D)) &\leq \psi(\mu(C))\\ \sum_{C\in\xi}\sum_{D\subseteq C}\psi(\mu(D)) &\leq \sum_{C\in\xi}\psi(\mu(C))\\ H_\mu(\xi) = -\sum_{C\in\xi}\psi(\mu(C)) &\leq -\sum_{C\in\xi}\sum_{D\subseteq C}\psi(\mu(D))\\ &= -\sum_{D\in\eta}\psi(\mu(D)) = H_\mu(\eta).\end{aligned}$$

Por lo tanto,

$$H_\mu(\xi) \leq H_\mu(\eta).$$

□

4.1.3. Entropía Condicional

Antes de definir lo que entenderemos por "entropía condicional", vamos a definir primero una operación entre particiones medibles que denotaremos con el símbolo $\vee$. La operación de particiones medibles es la generalización del "espacio producto" en probabilidad cuando los esquemas son mutuamente independientes, es decir, dados dos esquemas finitos A y B con eventos $A_1, A_2, \cdots, A_n$ y $B_1, B_2, \cdots, B_m$ respectivamente, se puede construir un tercer esquema finito AB cuyos eventos A_iBj consisten en que ocurran al mismo tiempo los eventos A_i y B_j. Sean ξ, η particiones medibles de X. Definimos,

$$\xi \vee \eta := \{C \cap D : C \in \xi, D \in \eta\}. \tag{4.2}$$

Algunas propiedades inmediatas son las siguientes. Si ξ, η son particiones medibles de X, entonces tenemos,

- $\xi \vee \xi = \xi$.
- $\xi \vee \eta = \eta \vee \xi$.

La siguiente proposición muestra que la nueva operación entre particiones medibles está bien definida, es decir, es a su vez una partición medible.

Proposición 32. *Sea (X, S, μ) un espacio de medida y ξ, η particiones medibles. Entonces $\xi \vee \eta$ es una partición medible.*

Demostración. Veamos que se cumple que $\mu(\bigcup_{H \in \xi \vee \eta} H) = 1$. Por definición se tiene,

$$\bigcup_{\xi \vee \eta} H = \bigcup_{C \in \xi D \in \eta} C \cap D = \bigcup_{C \in \xi} C \cap \bigcup_{D \in \eta} D.$$

Observamos que,

$$\begin{aligned} X \setminus \bigcup_{\xi \vee \eta} H &= X \setminus \left(\bigcup_{C \in \xi} C \cap \bigcup_{D \in \eta} D \right) \\ &= X \setminus \bigcup_{C \in \xi} C \cup X \setminus \bigcup_{D \in \eta} D. \end{aligned}$$

Esto implica,

$$\begin{aligned} 0 \leq \mu \left(X \setminus \bigcup_{\xi \vee \eta} H \right) &\leq \mu \left(X \setminus \bigcup_{C \in \xi} C \right) + \mu \left(X \setminus \bigcup_{D \in \eta} D \right) \\ &= 0. \end{aligned}$$

Entonces $\mu(X \setminus \bigcup_{\xi \vee \eta} H) = 0$ y por lo tanto $\mu(\bigcup_{\xi \vee \eta} H) = 1$. Ahora, supongamos que $H, K \in \xi \vee \eta$ y veamos que se cumple que $\mu(H \cap K) = 0$. Por definición,

$$\begin{aligned} H &= C \cap D \qquad C \in \xi, D \in \eta \\ K &= C' \cap D' \qquad C' \in \xi, D' \in \eta. \end{aligned}$$

Luego,

$$0 \leq \mu(H \cap K) \leq \mu(C \cap D \cap C' \cap D') \leq \mu(C \cap C') = 0.$$

Por lo tanto $\mu(H \cap K) = 0$. Así pues, $\xi \vee \eta$ es una partición medible. □

El siguiente lema muestra la relación entre la entropía de dos particiones ξ, η y la partición $\xi \vee \eta$.

Lema 11. *Sea (X, S, μ) un espacio de medida y sean ξ y η particiones medibles de X. Entonces,*

$$H_\mu(\xi \vee \eta) \leq H_\mu(\xi) + H_\mu(\eta).$$

Demostración. Por definición,

$$\begin{aligned} H_\mu(\xi \vee \eta) &= - \sum_{C \in \xi, D \in \eta} \mu(C \cap D) \log(\mu(C \cap D)) \\ &= - \sum_{C \in \xi, D \in \eta} \mu(C \cap D) \log \left(\frac{\mu(C \cap D)}{\mu(C)} \mu(C) \right). \end{aligned}$$

Pero,

$$\log\left(\frac{\mu(C\cap D)}{\mu(C)}\mu(C)\right) = \log\left(\frac{\mu(C\cap D)}{\mu(C)}\right) + \log(\mu(C)).$$

Entonces tenemos,

$$\begin{aligned} H_\mu(\xi\vee\eta) &= -\sum_{C\in\xi, D\in\eta}\left(\mu(C\cap D)\log\left(\frac{\mu(C\cap D)}{\mu(C)}\right) + \mu(C\cap D)\log(\mu(C))\right) \\ &= -\sum_{C\in\xi, D\in\eta}\mu(C)\frac{\mu(C\cap D)}{\mu(C)}\log\left(\frac{\mu(C\cap D)}{\mu(C)}\right) \\ &-\sum_{C\in\xi, D\in\eta}\mu(C\cap D)\log(\mu(C))). \end{aligned}$$

Así también, como,

$$\psi\left(\frac{\mu(C\cap D)}{\mu(C)}\right) = \frac{\mu(C\cap D)}{\mu(C)}\log\left(\frac{\mu(C\cap D)}{\mu(C)}\right).$$

Esto implica que,

$$H_\mu(\xi\vee\eta) = -\sum_{C\in\xi, D\in\eta}\mu(C)\psi\left(\frac{\mu(C\cap D)}{\mu(C)}\right) - \sum_{C\in\xi, D\in\eta}\mu(C\cap D)\log(\mu(C))).$$

Luego, como $\sum_{C\in\xi}\mu(C) = \mu(\bigcup_{C\in\xi}C) = 1$. Por la desigualdad de Jensen aplicada al primer sumando de la igualdad anterior,

$$\psi\left(\sum_{C\in\xi}\mu(C)\frac{\mu(C\cap D)}{\mu(C)}\right) \leq \sum_{C\in\xi}\mu(C)\psi\left(\frac{\mu(C\cap D)}{\mu(C)}\right)$$

$$-\sum_{C\in\xi}\mu(C)\psi\left(\frac{\mu(C\cap D)}{\mu(C)}\right) \leq -\psi\left(\sum_{C\in\xi}\mu(C\cap D)\right).$$

Así pues,

$$H_\mu(\xi\vee\eta) \leq -\sum_{D\in\eta}\psi\left(\sum_{C\in\xi}\mu(C\cap D)\right) - \sum_{C\in\xi, D\in\eta}\mu(C\cap D)\log(\mu(C)).$$

Como,

$$\textstyle\sum_{D\in\eta}\mu(D\cap C) = \mu(C) \quad \text{y} \quad \sum_{C\in\xi}\mu(C\cap D) = \mu(D).$$

Entonces,

$$-\sum_{D\in\eta}\psi\left(\sum_{C\in\xi}\mu(C\cap D)\right) = -\sum_{D\in\eta}\psi(\mu(D)) = H_\mu(\eta).$$

También,

$$-\sum_{C\in\xi,D\in\eta}\mu(C\cap D)\log(\mu(C))) = -\sum_{C\in\xi}\mu(C)\log(\mu(C))) = H_\mu(\xi).$$

Por lo tanto,

$$H_\mu(\xi\vee\eta)\leq H_\mu(\xi)+H_\mu(\eta).$$

□

En lo que sigue, definiremos la entropía condicional. El espacio producto de dos esquemas finitos A y B consistía en eventos A_iB_j, es decir, eventos en los que ocurre al mismo tiempo, tanto el evento A_i, como el evento B_j. Siempre y cuando los eventos A_i y B_j fueran eventos mutuamente independientes. Vamos a suponer que que los eventos de los esquemas finitos son mutuamente dependientes, es decir, nos interesa la probabilidad de que ocurra un evento B_j en el esquema B, dado que ocurrió un evento A_i en el esquema A. En Probabilidad se define la *probabilidad condicional* como,

$$P_{A_i}(B_j) = \frac{P(A_iB_j)}{P(A_i)}.$$

Al calcular la entropía con esta nueva probabilidad obtenemos,

$$H_{A_i}(B) = -\sum_{k=1}^{m}P_{A_i}(B_k)\log(P_{A_i}(B_k)).$$

El valor H_{A_i} corresponde a cada evento A_i del esquema finito A. El valor $H_{A_i}(B)$ se puede ver como una variable aleatoria definida sobre el espacio A, de manera que la esperanza matemática de esta variables es la entropía condicional del espacio B sobre el espacio A.

$$H_A(B) = -\sum_{i=1}^{m}P(A_i)H_{A_i}(B).$$

La cantidad $H_A(B)$ indica la cantidad de información que está contenida en el espacio B, siempre y cuando sepamos cuál de los eventos en A ocurran. A continuación vamos a generalizar esta idea para un espacio de medida.

Definición 27. *Sea (X,S,μ) un espacio de medida. Sean ξ,ζ dos particiones medibles de X. Definimos la **entropía condicional** de ξ con respecto a ζ como:*

$$H_\mu(\xi|\zeta) = \sum_{C\in\xi,D\in\zeta}\mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)}.$$

Vamos a probar algunas de las propiedades más relevantes que tiene la entropía condicional, para ello, definiremos un conjunto auxiliar. Sea ξ una partición medible de X y $D\subseteq X$ un conjunto medible. Definimos,

$$\xi|D = \{C\cap D : C\in\xi\}.$$

Proposición 33. *Sea (X, S, μ) un espacio de medida, ξ una partición medible de X y $D \subseteq X$ un conjunto medible. Entonces el conjunto $\xi|D$ es una partición medible de D.*

Demostración. Para ver que $\xi|D$ es una partición medible, debemos verificar,

- $\mu(\bigcup_{C\in\xi|D} C) = \mu(D)$.
- $\mu(C \cap H) = 0$ para $C, H \in \xi|D$, $C \neq H$.

Veamos que se cumple la primera propiedad. Por hipótesis sabemos que ξ es una partición medible. Luego,

$$\mu\left(\bigcup_{C\in\xi|D} C\right) = \mu\left(\bigcup_{C\in\xi} C \cap D\right) = \mu(D).$$

Para ver que se cumple la segunda propiedad, consideremos $C, H \in \xi|D$ tales que $C \neq H$ y veamos que $\mu(C \cap H) = 0$.

$$C = C' \cap D \qquad \text{con } C' \in \xi$$

$$H = H' \cap D \qquad \text{con } H' \in \xi.$$

Entonces, como ξ es partición medible de X, se tiene,

$$0 \leq \mu(C \cap H) = \mu(C' \cap D \cap H' \cap D) = \mu(C' \cap H' \cap D) \leq \mu(C' \cap H') = 0.$$

Por lo tanto, $\xi|D$ es partición medible de D. □

Dada una medida μ y $D \subseteq X$ un conjunto medible, podemos definir lo que llamaremos su "medida condicional" como sigue,

$$\mu(C|D) = \mu_D(C) = \frac{\mu(C \cap D)}{\mu(D)}.$$

Más aún, observamos que,

$$\begin{aligned} H_{\mu_D}(\xi|D) &= -\sum_{C\in\xi} \mu_D(C)\log(\mu_D(C)) \\ &= -\sum_{C\in\xi} \mu(C|D)\log(\mu(C|D)). \end{aligned} \tag{4.3}$$

Proposición 34. *Sea (X, S, μ) un espacio de medida y ξ, ζ particiones medibles de X. Entonces,*

$$H_\mu(\xi|\zeta) = \sum_{D\in\zeta} \mu(D) H_{\mu_D}(\xi|D).$$

Demostración. Por definición,

$$\begin{aligned}
H_\mu(\xi|\zeta) &= -\sum_{C\in\xi, D\in\zeta} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi, D\in\zeta} \mu(D)\frac{\mu(C\cap D)}{\mu(D)}\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi, D\in\zeta} \mu(D)\psi(\mu(C|D)) \\
&= -\sum_{D\in\zeta} \mu(D)\sum_{C\in\xi}\psi(\mu(C|D)) \\
&= -\sum_{D\in\zeta} \mu(D)H_{\mu_D}(\xi|D).
\end{aligned}$$

□

El siguiente resultado nos muestra la relación que existe entre la entropía condicional y los refinamientos de las particiones medibles.

Lema 12. *Sea (X, S, μ) un espacio de medida y ζ, ξ, η particiones medibles de X. Si ζ es un refinamiento de η, entonces se tiene,*

$$H_\mu(\xi|\zeta) \le H_\mu(\xi|\eta).$$

Demostración. Por definición,

$$\begin{aligned}
H_\mu(\xi|\zeta) &= -\sum_{C\in\xi, D\in\zeta} \mu(C\cap D)\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi, D\in\zeta} \mu(D)\frac{\mu(C\cap D)}{\mu(D)}\log\frac{\mu(C\cap D)}{\mu(D)}.
\end{aligned}$$

Como $\mu(D) = \sum_{E\in\eta}\mu(E\cap D)$, entonces,

$$\begin{aligned}
H_\mu(\xi|\zeta) &= -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(E)\frac{\mu(E\cap D)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)}\log\frac{\mu(C\cap D)}{\mu(D)} \\
&= -\sum_{C\in\xi, E\in\eta} \mu(E)\sum_{D\in\zeta}\frac{\mu(E\cap D)}{\mu(E)}\psi\left(\frac{\mu(C\cap D)}{\mu(D)}\right).
\end{aligned}$$

Como,

$$\sum_{D\in\zeta}\frac{\mu(D\cap E)}{\mu(E)} = \frac{\sum_{D\in\zeta}\mu(D\cap E)}{\mu(E)} = \frac{\mu(E)}{\mu(E)} = 1.$$

Puesto que sabemos que ψ es una función convexa, por la desigualdad de Jensen tenemos,

$$\psi\left(\sum_{D\in\zeta}\frac{\mu(E\cap D)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)}\right) \le \sum_{D\in\zeta}\frac{\mu(E\cap D)}{\mu(E)}\psi\left(\frac{\mu(C\cap D)}{\mu(D)}\right).$$

Así pues,

$$H_\mu(\xi|\zeta) \leq -\sum_{C\in\xi, E\in\eta} \mu(E)\psi\left(\sum_{D\in\zeta} \frac{\mu(E\cap D)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)}\right).$$

Como por hipótesis tenemos que ζ es refinamiento de η, entonces por definición sabemos que $\forall E \in \zeta$, $\exists D \in \eta$ tal que $\mu(E\setminus D) = 0$. Esto implica que $\forall D \in \zeta$, $\exists E \in \eta$ tal que $\mu(D\cap E) = 0$ o bien, $\mu(D\cap E) = \mu(D)$. Entonces,

$$\sum_{D\in\zeta} \frac{\mu(D\cap E)}{\mu(E)}\frac{\mu(C\cap D)}{\mu(D)} = \sum_{D\in\zeta, D\subseteq E} \frac{\mu(C\cap E)}{\mu(E)}.$$

Sin embargo, como $E = \bigcup_{D\subseteq E} D \cup A$ donde $\mu(A) = 0$, tenemos,

$$C\cap E = C\cap \bigcup_{D\subseteq E} D\cup C\cap A \qquad \text{con } \mu(C\cap A)\leq \mu(A) = 0.$$

Entonces,

$$\mu(C\cap E) = \mu\left(C\cap \bigcup_{D\subseteq E} D\right) = \sum_{D\subseteq E, D\in\zeta} \mu(C\cap D).$$

Luego pues,

$$\sum_{D\in\zeta, D\subseteq E} \frac{\mu(C\cap E)}{\mu(E)} = \frac{\mu(C\cap E)}{\mu(E)}.$$

Es decir,

$$\begin{aligned} H_\mu(\xi|\zeta) &\leq -\sum_{C\in\xi, E\in\eta} \mu(E)\psi\left(\frac{\mu(C\cap E)}{\mu(E)}\right) \\ &= -\sum_{C\in\xi, E\in\eta} \mu(E)\frac{\mu(C\cap E)}{\mu(E)}\log\left(\frac{\mu(C\cap E)}{\mu(E)}\right) \\ &= H_\mu(\xi|\eta). \end{aligned}$$

Por lo tanto,

$$H_\mu(\xi|\zeta) \leq H_\mu(\xi|\eta).$$

□

Finalmente presentamos el resultado principal que muestra la relación que existe entre la entropía de la operación entre particiones que definimos anteriormente y la la entropía condicional.

Lema 13. *Sea (X, S, μ) un espacio de medida y sean ξ, ζ, η particiones medibles de X, entonces,*

$$H_\mu(\xi\vee\zeta|\eta) = H_\mu(\xi|\zeta\vee\eta) + H_\mu(\zeta|\eta).$$

Demostración. Por definición,

$$\begin{aligned} H_\mu(\xi \vee \zeta|\eta) &= -\sum_{H\in\xi\vee\zeta, E\in\eta} \mu(H\cap E)\log\frac{\mu(H\cap E)}{\mu(E)} \\ &= -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(C\cap D\cap E)}{\mu(E)}. \end{aligned}$$

Por propiedades del logaritmo,

$$\begin{aligned} H_\mu(\xi \vee \zeta|\eta) = &-\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(C\cap D\cap E)}{\mu(D\cap E)} \\ &-\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(D\cap E)}{\mu(E)}. \end{aligned}$$

Pero, por definición, tenemos,

$$H_\mu(\xi|\zeta\vee\eta) = -\sum_{C\in\xi, D\in\zeta, E\in\eta} \mu(C\cap D\cap E)\log\frac{\mu(C\cap D\cap E)}{\mu(D\cap E)}.$$

Observamos,

$$\mu(D\cap E) = \sum_{C\in\xi}\mu(C\cap D\cap E).$$

Por lo tanto,

$$\begin{aligned} H_\mu(\xi \vee \zeta|\eta) &= H_\mu(\xi|\zeta\vee\eta) - \sum_{D\in\zeta, E\in\eta} \mu(D\cap E)\log\frac{\mu(D\cap E)}{\mu(E)} \\ &= H_\mu(\xi|\zeta\vee\eta) + H_\mu(\zeta|\eta). \end{aligned}$$

□

El resultado que acabamos de probar generaliza el Lema 11. Consideremos la partición trivial $\eta = \{X\}$. Entonces se tiene,

$$\xi\vee\zeta|\eta = \xi\vee\zeta$$

$$\zeta\vee\eta = \zeta$$

$$\zeta|\eta = \zeta.$$

Es decir, en caso de que η fuera la partición trivial, por el Lema 13,

$$H_\mu(\xi\vee\zeta) = H_\mu(\xi|\zeta) + H_\mu(\zeta). \tag{4.4}$$

Observamos que cuando los eventos son mutuamente independientes se tiene,

$$H_\mu(\xi|\zeta) = H_\mu(\xi).$$

Por lo que la igualdad 4.4 se convierte en,

$$H_\mu(\xi \vee \zeta) = H_\mu(\xi) + H_\mu(\zeta).$$

La segunda observación que se puede hacer es que de la ecuación (4.4) y del Lema 11 se sigue,

$$H_\mu(\xi|\zeta) \leq H_\mu(\xi).$$

Esto significa que la cantidad de información de una partición medible ξ, sólo puede decrecer o permanecer igual si se condiciona a la ocurrencia de un elemento de la partición ζ.

4.2. Entropía de una Transformación y una Partición

En esta sección analizaremos lo que le ocurre a la entropía de una partición al introducir una transformación medible T y sus iteraciones. Probaremos algunos resultados auxiliares para dar una primera caracterización de la entropía de una partición medible con respecto a una transformación en un espacio X.

Consideremos una partición medible ξ de un espacio X y una transformación medible $T : X \to X$. Definimos el siguiente conjunto,

$$T^{-1}(\xi) := \{T^{-1}(C) : C \in \xi\}.$$

A continuación veremos que $T^{-1}(\xi)$ es a su vez una partición medible. Antes de probar el siguiente resultado, notamos que una propiedad inmediata que se sigue de la definición de la imagen inversa de una partición medible y que utilizaremos en la prueba de la siguiente proposición, es que,

$$T^{-1}(\xi \vee \eta) = T^{-1}(\xi) \vee T^{-1}(\eta).$$

Proposición 35. *Sean (X, S, μ) un espacio medible y $T : X \to X$ una transformación que preserva la medida μ. Si ξ es una partición medible de X, entonces $T^{-1}(\xi)$ es una partición medible.*

Demostración. Veamos primero que $\mu(\bigcup_{c\in\xi} T^{-1}(C)) = 1$. Por propiedades de la imagen inversa de una función, se tiene que,

$$\mu\left(\bigcup_{c\in\xi} T^{-1}(C)\right) = \mu\left(T^{-1}\left(\bigcup_{c\in\xi} C\right)\right).$$

Por hipótesis, T es una transformación que preserva la medida μ. Entonces,

$$\mu\left(\bigcup_{c\in\xi} T^{-1}(C)\right) = \mu\left(\bigcup_{c\in\xi} C\right) = 1.$$

Ahora bien, sean $D, C \in \xi$. Veamos que $\mu(T^{-1}(D) \cap T^{-1}(C)) = 0$. Por propiedades de la imagen inversa, se tiene,

$$\mu(T^{-1}(D) \cap T^{-1}(C)) = \mu(T^{-1}(D \cap C)).$$

Como por hipótesis, la transformación T preserva la medida μ, entonces,

$$\mu(T^{-1}(D) \cap T^{-1}(C)) = \mu(D \cap C) = 0.$$

Por lo tanto, se concluye que $T^{-1}(\xi)$ es una partición medible. $\square$

La proposición anterior nos dice que tiene sentido hablar de la entropía de la imagen inversa de una partición medible. El siguiente lema muestra que en el caso de que la transformación preserve la medida, la entropía de una partición medible coincide con la entropía de la imagen inversa de una partición medible, es decir, no se añade ni disminuye la cantidad de información al sistema.

Lema 14. *Sean (X, S, μ) un espacio medible y $T : X \to X$ una transformación que preserva la medida μ. Si ξ es una partición medible de X, entonces,*

$$H_\mu(T^{-1}(\xi)) = H_\mu(\xi).$$

Demostración. Por definición de entropía, tenemos,

$$H_\mu(T^{-1}(\xi)) = -\sum_{C \in \xi} \mu(T^{-1}(C)) \log \mu(T^{-1}(C)).$$

Como por hipótesis la transformación T preserva la medida μ, entonces,

$$\begin{aligned} H_\mu(T^{-1}(\xi)) &= -\sum_{C \in \xi} \mu(C) \log \mu(C) \\ &= H_\mu(\xi). \end{aligned}$$

$\square$

Ahora bien, para una partición medible ξ de un espacio X y una transformación $T : X \to X$ que preserva una medida μ, definimos:

$$\xi_n = \bigvee_{k=0}^{n-1} T^{-k}(\xi) \qquad \forall n \in \mathbb{N}.$$

Las siguientes dos propiedades son propiedades auxiliares que nos ayudarán a definir la entropía de una transformación que preserva la medida.

Proposición 36. *Sea (X, S, μ) un espacio de medida, $T : X \to X$ una transformación que preserva la medida μ y ξ es una partición medible de X. Entonces para $n, m \in \mathbb{N}$ se tiene,*

$$H_\mu(\xi_{n+m}) \le H_\mu(\xi_n) + H_\mu(\xi_m).$$

Demostración. Por definición y por los lemas 11 y 14 se tiene que,

$$\begin{aligned} H_\mu(\xi_{n+m}) &= H_\mu(\xi_n \vee T^{-n}(\xi_m)) \\ &\leq H_\mu(\xi_n) + H_\mu(T^{-n}(\xi_m)) \\ &= H_\mu(\xi_n) + H_\mu(\xi_m). \end{aligned}$$

Por lo tanto,

$$H_\mu(\xi_{n+m}) \leq H_\mu(\xi_n) + H_\mu(\xi_m).$$

□

El siguiente lema es un resultado sobre sucesiones de números reales positivos y nos servirá para asegurar que la entropía de una transformación con respecto a una partición medible esta bien definida.

Lema 15. *Si $(a_n)_{n\in\mathbb{N}}$ es una sucesión de números reales positivos tales que $a_{n+m} \leq a_n + a_m$ $\forall n, m \in \mathbb{N}$, entonces el siguiente límite:*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf\left\{\frac{a_k}{k} : k \in \mathbb{N}\right\}$$

existe.

Demostración. Sea $k \in \mathbb{N}$. Por el algoritmo de la división, tenemos que existe $q \in \mathbb{N}$ y $r \in \{0, 1, \cdots, k-1\}$ tales que $n = qk + r$. Por hipótesis,

$$\frac{a_n}{n} = \frac{a_{qk+r}}{qk+r} \leq \frac{a_{qk} + a_r}{qk+r}.$$

Así también, por hipótesis,

$$a_{qk} \leq q a_k.$$

Esto implica lo siguiente,

$$\frac{a_n}{n} \leq \frac{a_{qk} + a_r}{qk+r} \leq \frac{q a_k + a_r}{qk+r}.$$

Tomamos el límite superior,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \limsup_{q\to\infty} \frac{q(a_k + \frac{a_r}{q})}{q(k + \frac{r}{q})} = \lim_{q\to\infty} \frac{a_k + \frac{a_r}{q}}{k + \frac{r}{q}} = \frac{a_k}{k}.$$

Es decir,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_k}{k} \qquad \forall k \in \mathbb{N}.$$

Entonces,

$$\limsup_{n\to\infty} \frac{a_n}{n} \leq \inf\left\{\frac{a_k}{k} : k \in \mathbb{N}\right\} \leq \sup_{m>1} \inf_{n>m} \frac{a_k}{k} = \liminf_{n\to\infty} \frac{a_n}{n}.$$

Como siempre se tiene que,

$$\liminf_{n\to\infty} \frac{a_n}{n} \leq \limsup_{n\to\infty} \frac{a_n}{n}.$$

Concluimos que el límite existe y además,

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf\left\{\frac{a_k}{k} : k \in \mathbb{N}\right\}.$$

□

El siguiente teorema utiliza los resultados anteriores para probar una importante propiedad que nos permitirá definir a la entropía de una transformación con respecto a una partición medible.

Teorema 27. *Sea (X, S, μ) un espacio medible tal que $\mu(X) = 1$ y $T : X \to X$ una trasformación medible que preserva la medida μ. Sea ξ una partición medible de X, entonces el límite,*

$$\lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n) = \inf_{n\in\mathbb{N}} \frac{1}{n} H_\mu(\xi_n) \tag{4.5}$$

existe.

Demostración. Por la Proposición 36 se tiene que,

$$H_\mu(\xi_{n+m}) \leq H_\mu(\xi_n) + H_\mu(\xi_m) \qquad \forall n, m \in \mathbb{N}.$$

Por el Lema 15 concluimos,

$$\lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n) = \inf_{n\in\mathbb{N}} \frac{1}{n} H_\mu(\xi_n).$$

□

Ahora bien, la entropía de una transformación con respecto a una partición medible es el ínfimo de los valores de $\frac{1}{n}H(\xi_n)$. El Teorema 27, afirma que ese valor es igual al límite cuando n tiende a infinito. A continuación formalizamos la definición.

Definición 28. *Sea (X, S, μ) un espacio medible tal que $\mu(X) = 1$ y $T : X \to X$ una trasformación medible que preserva la medida μ. Definimos la* ***entropía de T con respecto a μ y a una partición medible*** *ξ de X como sigue:*

$$h_\mu(T, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n).$$

La entropía de una transformación con respecto a una medida y a una partición medible se puede interpretar como la cantidad promedio de información por unidad de tiempo, o bien, por cada iteración de la transformación. En Teoría de la Información, este valor es importante porque representa la tasa bajo la cual se emite información, es decir, si consideramos una una fuente que emite un una sucesión de símbolos, lo que se está midiendo es la cantidad promedio dada por cada símbolo emitido. Así pues, es una manera de evaluar la efectividad de una fuente.

Otra manera útil y edificante de calcular la entropía de una transformación medible es a partir de la noción de entropía condicional. A continuación se presentan dos proposiciones auxiliares y posteriormente la caracterización.

Proposición 37. *Sea (X, S, μ) un espacio de medida, $T : X \to X$ una transformación medible y ξ una partición medible de X, entonces se tiene que,*

$$\bigvee_{i=1}^{n+1} T^{-i}(\xi),$$

es un refinamiento de,

$$\bigvee_{i=1}^{n} T^{-i}(\xi) \qquad \forall n \in \mathbb{N}.$$

Demostración. En general, para $n \in \mathbb{N}$ se tiene que,

$$\bigvee_{i=1}^{n+1} T^{-i}(\xi) = T^{-1}(\xi) \vee T^{-2}(\xi) \vee \cdots \vee T^{-(n+1)}(\xi)$$

$$\bigvee_{i=1}^{n} T^{-i}(\xi) = T^{-1}(\xi) \vee T^{-2}(\xi) \vee \cdots \vee T^{-n}(\xi).$$

Veamos que para cada $C \in \bigvee_{i=1}^{n+1} T^{-i}(\xi)$, existe $D \in \bigvee_{i=1}^{n} T^{-i}(\xi)$ tal que $\mu(C \setminus D) = 0$. Sea $C \in \bigvee_{i=1}^{n+1} T^{-i}(\xi)$. Entonces, por definición, C es de la forma,

$$C = C_1 \cap C_2 \cap \cdots \cap C_{n+1} \qquad \text{con } C_i \in T^{-i}(\xi).$$

Tomamos,

$$D = C_1 \cap C_2 \cap \cdots \cap C_n \qquad \text{con } C_i \in T^{-i}(\xi).$$

Entonces,

$$\begin{aligned}
\mu(C \setminus D) &= \mu((C_1 \cap C_2 \cap \cdots \cap C_{n+1}) \setminus (C_1 \cap C_2 \cap \cdots \cap C_n)) \\
&\leq \mu((C_1 \cap C_2 \cap \cdots \cap C_n) \setminus (C_1 \cap C_2 \cap \cdots \cap C_n)) \\
&= \mu(\emptyset) = 0.
\end{aligned}$$

□

La siguiente proposición afirma que la entropía condicional de una partición con respecto a la imagen inversa de la partición es no decreciente. Para probar el resultado ocupamos la proposición anterior.

Proposición 38. *Sea* (X, S, μ) *un espacio de medida,* $T : X \to X$ *una transformación medible y* ξ *una partición medible de* X*. Entonces,*

$$H_\mu \left(\xi \middle| \bigvee_{i=1}^{n} T^{-i}(\xi) \right)$$

es no creciente.

Demostración. Por la Proposición 37, tenemos que $\bigvee_{i=1}^{n+1} T^{-i}(\xi)$ es un refinamiento de $\bigvee_{i=1}^{n} T^{-i}(\xi)$. Por el Lema 12 tenemos que,

$$H_\mu \left(\xi \middle| \bigvee_{i=1}^{n+1} T^{-i}(\xi) \right) \leq H_\mu \left(\xi \middle| \bigvee_{i=1}^{n} T^{-i}(\xi) \right).$$

Por lo tanto, $H_\mu \left(\xi | \bigvee_{i=1}^{n} T^{-i}(\xi) \right)$ es no creciente. □

El siguiente teorema es una primera caracterización de la entropía de una transformación con respecto a una partición medible con base en la entropía condicional.

Teorema 28. *Sea* (X, S, μ) *un espacio de medida tal que* $\mu(X) = 1$*. Sea* $T : X \to X$ *una transformación medible que preserva la medida* μ *y* ξ *una partición medible de* X*, entonces,*

$$h_\mu(T, \xi) = \lim_{n \to \infty} H_\mu \left(\xi \middle| \bigvee_{i=1}^{n} T^{-i} \xi \right).$$

Demostración. Por la Proposición 38 sabemos que $H_\mu(\xi | \bigvee_{i=1}^{n} T^{-i}(\xi))$ es no creciente. Esto implica que

$$\lim_{n \to \infty} H_\mu \left(\xi \middle| \bigvee_{i=1}^{n} T^{-i}(\xi) \right).$$

existe. Por el Lema 13, sabemos que si ξ, ζ, η son particiones medibles de X, entonces se tiene,

$$H_\mu(\xi \vee \zeta | \eta) = H_\mu(\xi | \zeta \vee \eta) + H_\mu(\zeta | \eta).$$

Consideremos la partición trivial $\eta = \{X\}$. Entonces se tiene,

$$\xi \vee \zeta | \eta = \xi \vee \zeta$$

$$\zeta \vee \eta = \zeta$$

$$\zeta | \eta = \zeta.$$

Es decir, en caso de que η fuera la partición trivial, por el Lema 13 se tiene que,

$$H_\mu(\xi \vee \zeta) = H_\mu(\xi | \zeta) + H_\mu(\zeta). \tag{4.6}$$

Afirmación:

$$H_\mu(\xi_n) = H_\mu(\xi) + \sum_{j=1}^{n-1} H_\mu\left(\xi \middle| \bigvee_{i=1}^{j} T^{-i}(\xi)\right) \qquad \forall n \in \mathbb{N}.$$

Por inducción sobre n. Tomamos como base de la inducción a $n = 1$ en el cual se cumple la afirmación. En efecto, pues,

$$H_\mu(\xi_1) = H_\mu(\xi).$$

Hacemos la hipótesis de inducción. Supongamos que la afirmación se cumple para n, es decir, supongamos que ocurre,

$$H_\mu(\xi_n) = H_\mu(\xi) + \sum_{j=1}^{n-1} H_\mu\left(\xi \middle| \bigvee_{i=1}^{j} T^{-i}(\xi)\right).$$

Para verificar el paso inductivo, veamos que la afirmación se cumple para $n+1$. Por definición ocurre lo siguiente,

$$H_\mu(\xi_{n+1}) = H_\mu\left(\bigvee_{i=1}^{n} T^{-i}(\xi) \vee \xi\right).$$

Por la Observación (4.6) se sigue que,

$$H_\mu(\xi_{n+1}) = H_\mu\left(\bigvee_{i=1}^{n} T^{-i}(\xi)\right) + H_\mu\left(\xi \middle| \bigvee_{i=1}^{n} T^{-i}(\xi)\right).$$

Ahora bien, como,

$$\bigvee_{i=1}^{n} T^{-i}(\xi) = T^{-1} \bigvee_{i=0}^{n-1} T^{-i}(\xi).$$

Por el Lema 14 tenemos,

$$\begin{aligned} H_\mu\left(\bigvee_{i=1}^{n} T^{-i}(\xi)\right) &= H_\mu\left(T^{-1} \bigvee_{i=0}^{n-1} T^{-i}(\xi)\right) \\ &= H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\xi)\right) \\ &= H_\mu(\xi_n). \end{aligned}$$

Luego pues,

$$H_\mu(\xi_{n+1}) = H_\mu(\xi_n) + H_\mu\left(\xi \middle| \bigvee_{i=1}^{n} T^{-i}(\xi)\right).$$

Al sustituir la hipótesis de inducción en la ecuación anterior obtenemos,

$$H_\mu(\xi_{n+1}) = H_\mu(\xi) + \sum_{j=1}^{n} H_\mu\left(\xi \middle| \bigvee_{i=1}^{n} T^{-i}(\xi)\right).$$

Por lo tanto,

$$H_\mu(\xi_n) = H_\mu(\xi) + \sum_{j=1}^{n-1} H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{j} T^{-i}(\xi)\right) \qquad \forall n \in \mathbb{N}.$$

Con base en la definición y utilizando la la afirmación que acabamos de probar, podemos afirmar lo siguiente,

$$\begin{aligned} h_\mu(T,\xi) = &= \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n) \\ &= \lim_{n\to\infty} \frac{1}{n}\left(H_\mu(\xi) + \sum_{j=1}^{n-1} H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{j} T^{-i}(\xi)\right)\right) \\ &= \lim_{n\to\infty} \frac{1}{n}\left(\sum_{j=1}^{n-1} H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{j} T^{-i}(\xi)\right)\right) \\ &= \lim_{n\to\infty} H_\mu\left(\xi \,\middle|\, \bigvee_{i=1}^{j} T^{-i}(\xi)\right). \end{aligned}$$

□

El último teorema nos muestra que la entropía de una transformación respecto a una partición medible se puede interpretar como la información promedio añadida por el estado actual una vez que se conocen todos los estados anteriores.

4.3. Entropía de una Transformación

En esta sección definimos la ***entropía de Kolomogorov-Sinaí***, o bien, la ***entropía de una transformación*** T con respecto a una medida μ como sigue:

$$h_\mu(T) = \sup\{h_\mu(T,\xi) : \xi \text{ es una partición medible de } X\}.$$

La entropía de una transformación T respecto a una medida μ se puede interpretar como el valor más grande de la información promedio que se puede obtener bajo un número infinito de iteraciones de la transformación T. A continuación veremos una importante propiedad de la entropía de una transformación con respecto a una medida. Probaremos que la entropía de T^k es igual a k veces la entropía de T.

Proposición 39. *Sea (X, S, μ) un espacio medible tal que $\mu(X) = 1$ y $T : X \to X$ una trasformación medible que preserva la medida μ. Entonces*

$$h_\mu(T^k) = k h_\mu(T) \qquad \forall k \in \mathbb{N}.$$

Demostración. Sea ξ una partición medible de X. Entonces, por definición se tiene que,

$$h_\mu(T^k, \xi_k) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n),$$

donde,

$$\xi_n = \bigvee_{l=0}^{n-1} (T^k)^{-l}(\xi_k) = \bigvee_{l=0}^{n-1} T^{-kl}(\xi_k).$$

Sin embargo, tenemos que,

$$\begin{aligned}
\bigvee_{l=0}^{n-1} T^{-kl}(\xi_k) &= \bigvee_{l=0}^{n-1} T^{-kl} \left(\bigvee_{m=0}^{k-1} T^{-m}(\xi) \right) \\
&= \bigvee_{l=0}^{n-1} \left(\bigvee_{m=0}^{k-1} T^{-kl-m}(\xi) \right) \\
&= \bigvee_{i=0}^{nk-1} T^{-i}(\xi).
\end{aligned}$$

Así pues,

$$\begin{aligned}
h_\mu(T^k, \xi_k) &= \lim_{n\to\infty} \frac{1}{n} H_\mu \left(\bigvee_{l=0}^{n-1} T^{-kl}(\xi_k) \right) \\
&= \lim_{n\to\infty} \frac{1}{n} H_\mu \left(\bigvee_{i=0}^{nk-1} T^{-i}(\xi) \right) \\
&= \lim_{n\to\infty} \frac{k}{kn} H_\mu \left(\bigvee_{i=0}^{nk-1} T^{-i}(\xi) \right) \\
&= k h_\mu(T, \xi).
\end{aligned}$$

Tomando supremos, obtenemos,

$$h_\mu(T^k) \geq \sup\{h_\mu(T^k, \xi_k) : \xi \text{ es partición medible}\} = k h_\mu(T).$$

Para obtener la otra desigualdad, observamos que,

$$\bigvee_{i=0}^{n-1} T^{-ki}\xi \leq \bigvee_{i=0}^{nk-1} T^{-i}\xi,$$

Por la Proposición 31 se sigue que,

$$H_\mu \left(\bigvee_{i=0}^{n-1} T^{-ki}\xi \right) \leq H_\mu \left(\bigvee_{i=0}^{nk-1} T^{-i}\xi \right).$$

Entonces,

$$\begin{aligned} h_\mu(T^k,\xi) &\leq \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{nk-1} T^{-i}\xi\right) \\ &= \lim_{n\to\infty} \frac{k}{kn} H_\mu\left(\bigvee_{i=0}^{nk-1} T^{-i}\xi\right) \\ &= k h_\mu(T,\xi). \end{aligned}$$

Tomando supremos,

$$h_\mu(T^k) \leq k h_\mu(T).$$

Por lo tanto, concluimos que $h_\mu(T^k) = k h_\mu(T)$. □

En lo que sigue vamos a ver que a partir de la noción de entropía condicional, es suficiente considerar sucesiones crecientes de particiones que generan la σ-álgebra de X. Dado un espacio de medida (X,S,μ) y una partición medible de X, denotamos $S(\xi)$ a la σ-álgebra generada por ξ, es decir, a la σ-álgebra más pequeña que contiene a los elementos de ξ. Observamos que,

$$\bigvee_{n\in I} S(\xi_n),$$

es una σ-álgebra. Más aún, cuando I es finito, tenemos que,

$$S\left(\bigvee_{n\in I} \xi_n\right) = \bigvee_{n\in I} S(\xi_n).$$

A continuación vamos a desarrollar una serie de resultados para probar un teorema que nos facilite calcular la entropía de una transformación. El primer resultado nos dice cómo se relaciona la entropía condicional de las particiones medibles.

Lema 16. *Sea (X,S,μ) un espacio de medida y sean α,β,γ particiones medibles de X. Entonces,*

$$H_\mu(\alpha\vee\beta|\gamma) \leq H_\mu(\alpha|\gamma) + H_\mu(\beta|\gamma).$$

Demostración. Por el Lema 13 se tiene que,

$$H_\mu(\alpha\vee\beta|\gamma) = H_\mu(\alpha|\beta\vee\gamma) + H_\mu(\beta|\gamma).$$

Como $\beta\vee\gamma$ es un refinamiento de γ, por el Lema 12,

$$H_\mu(\alpha|\beta\vee\gamma) \leq H_\mu(\alpha|\gamma).$$

Por lo tanto,

$$H_\mu(\alpha\vee\beta|\gamma) \leq H_\mu(\alpha|\gamma) + H_\mu(\beta|\gamma).$$

□

El siguiente resultado nos dice cómo se relaciona la entropía de una transformación respecto a diferentes particiones medibles.

Lema 17. *Sea (X, S, μ) un espacio de medida, $T : X \to X$ una transformación medible y ζ, η dos particiones medibles de X, entonces,*

$$h_\mu(T, \eta) \leq h_\mu(T, \zeta) + H_\mu(\eta|\zeta).$$

Demostración. Primero observamos que $\eta_n \vee \zeta_n$ es un refinamiento de η_n, pues si $E \in \eta_n \vee \zeta_n$, entonces es de la forma $E = C \cap D$ con $C \in \eta_n$ y $D \in \zeta_n$. Tomando $C \in \eta_n$, se tiene que,

$$\mu(E \setminus C) = \mu(C \cap D \setminus C) \leq \mu(C \setminus C) = \mu(\emptyset) = 0.$$

Por la Proposición 31 se tiene,

$$H_\mu(\eta_n) \leq H_\mu(\eta_n \vee \zeta_n).$$

Por el Lema 13 y tomando la última partición como la partición trivial se tiene,

$$H_\mu(\eta_n \vee \zeta_n) = H_\mu(\eta_n|\zeta_n) + H_\mu(\zeta_n).$$

Así pues,

$$H_\mu(\eta_n) \leq H_\mu(\eta_n|\zeta_n) + H_\mu(\zeta_n).$$

Como,

$$\begin{aligned}\eta_n &= \bigvee_{k=0}^{n-1} T^{-k}(\eta) \\ &= \eta \vee T^{-1}(\eta) \vee \cdots \vee T^{-(n-1)}(\eta).\end{aligned}$$

Por el Lema anterior 16 se sigue,

$$H_\mu(\eta_n|\zeta_n) \leq \sum_{i=0}^{n-1} H_\mu(T^{-i}(\eta)|\zeta_n).$$

Luego pues,

$$H_\mu(\eta_n) \leq H_\mu(\zeta_n) + \sum_{i=0}^{n-1} H_\mu(T^{-i}(\eta)|\zeta_n).$$

Como ζ_n es un refinamiento de $T^{-i}(\zeta)$ para todo $i \in \{1, 2, \cdots, n\}$, por el Lema 12 se obtiene,

$$H_\mu(T^{-i}(\eta)|\zeta_n) \leq H_\mu(T^{-i}(\eta)|T^{-i}(\zeta)) \qquad \forall i \in \{1, 2, \cdots, n\}.$$

Esto implica que,

$$\sum_{i=0}^{n-1} H_\mu(T^{-i}(\eta)|\zeta_n) \leq \sum_{i=0}^{n-1} H_\mu(T^{-i}(\eta)|T^{-i}(\zeta)).$$

Entonces,

$$H_\mu(\eta_n) \leq H_\mu(\zeta_n) + \sum_{i=0}^{n-1} H_\mu(T^{-i}(\eta)|T^{-i}(\zeta)).$$

Por definición y como T preserva la medida μ.

$$\begin{aligned}
H_\mu(T^{-i}(\eta)|T^{-i}(\zeta)) &= \sum_{E\in\eta, D\in\zeta} \mu(T^{-i}(E)\cap T^{-i}(D)) \log \frac{\mu(T^{-i}(E)\cap T^{-i}(D))}{\mu(T^{-i}(D))} \\
&= \sum_{E\in\eta, D\in\zeta} \mu(T^{-i}(E\cap D)) \log \frac{\mu(T^{-i}(E\cap D))}{\mu(T^{-i}(D))} \\
&= \sum_{E\in\eta, D\in\zeta} \mu(E\cap D) \log \frac{\mu(E\cap D)}{\mu(T^{-i}(D))} \\
&= H_\mu(\eta|\zeta).
\end{aligned}$$

Así pues,

$$H_\mu(\eta_n) \leq H_\mu(\zeta_n) + \sum_{i=0}^{n-1} H_\mu(\eta|\zeta) = H_\mu(\zeta_n) + nH_\mu(\eta|\zeta).$$

Esto implica que,

$$\frac{1}{n}H_\mu(\eta_n) \leq \frac{1}{n}H_\mu(\zeta_n) + H_\mu(\eta|\zeta).$$

Tomando límite cuando n tiende a infinito,

$$\lim_{n\to\infty} \frac{1}{n}H_\mu(\eta_n) \leq \lim_{n\to\infty} \frac{1}{n}H_\mu(\zeta_n) + H_\mu(\eta|\zeta).$$

Por lo tanto,

$$h_\mu(T,\eta) \leq h_\mu(T,\zeta) + H_\mu(\eta|\zeta).$$

□

El siguiente resultado nos indica un comportamiento que tiene la entropía condicional con respecto a las particiones medibles. Cuando se tiene una sucesión de particiones medibles tales que se van refinando progresivamente, si la σ-álgebra del espacio medible coincide con las intersecciones de los elementos de la σ-álgebra generada por cada elemento de la sucesión de particiones, entonces la entropía condicional de una partición medible tiende a cero.

Lema 18. *Sea (X, S, μ) un espacio de medida y η una partición medible de X. Si $(\alpha_n)_{n\in\mathbb{N}}$ es una sucesión de particiones medibles con*

$$\bigvee_{n=1}^{\infty} S(\alpha_n) = S,$$

tales que α_{n+1} es un refinamiento de α_n para todo $n \in \mathbb{N}$, entonces,

$$H_\mu(\eta|\alpha_n) \to 0 \qquad \text{si } n \to \infty.$$

Demostración. Supongamos que $\eta = \{C_1, C_2, \cdots, C_k\}$. Para cada $n \in \mathbb{N}$ e $i \in \{1, 2, \cdots, k\}$ vamos a suponer que $D_i^n \subseteq C_i$ es un conjunto en $S(\alpha_n)$ tal que,

$$\mu(C_i \setminus D_i^n) \to 0 \qquad \text{si } n \to \infty.$$

Definimos,

$$\beta_n = \{D_0^n, D_1^n, \cdots, D_k^n\},$$

donde $D_0^n = X \setminus \bigcup_{i=1}^k D_i^n$. Observamos que β_n es una partición medible y más aún, α_n la refina para cada $n \in \mathbb{N}$. Por definición se D_0^n se sigue que,

$$\mu(C_i \cap D_0^n) \to 0 \qquad \text{si } n \to \infty. \tag{4.7}$$

Como $C_i \in \eta$ y $D_i^n \subseteq C_i$, se deducen las siguientes propiedades:

- $C_i \cap D_i^n = D_i^n \qquad \forall i \in \{1, \cdots, k\}.$
- $C_i \cap D_j^n = \emptyset \qquad j \neq i.$

Así pues, se verifica lo siguiente,

$$\begin{aligned} H_\mu(\eta|\beta_n) &= -\sum_{i=1}^k \mu(C_i \cap D_i^n) \log \frac{\mu(C_i \cap D_i^n)}{\mu(D_i^n)} \\ &\quad - \sum_{i=1}^k \mu(C_i \cap D_0^n) \log \frac{\mu(C_i \cap D_0^n)}{\mu(D_0^n)} \\ &\quad - \sum_{j=1}^k \sum_{i \neq j} \mu(C_i \cap D_j^n) \log \frac{\mu(C_i \cap D_j^n)}{\mu(D_j^n)} \\ &= -\sum_{i=1}^k \mu(C_i \cap D_0^n) \log \frac{\mu(C_i \cap D_0^n)}{\mu(D_0^n)}. \end{aligned}$$

Por (4.7) se sigue,

$$H_\mu(\eta|\beta_n) = -\sum_{i=1}^k \mu(C_i \cap D_0^n) \log \frac{\mu(C_i \cap D_0^n)}{\mu(D_0^n)} \to 0 \qquad n \to \infty.$$

Como α_n es un refinamiento de β_n. Por el Lema 12 se tiene que,

$$H_\mu(\eta|\alpha_n) \leq H_\mu(\eta|\beta_n) \to 0 \qquad \text{si } n \to \infty.$$

Por lo tanto,

$$H_\mu(\eta|\alpha_n) \to 0 \qquad \text{si } n \to \infty.$$

□

Bajo las hipótesis del lema anterior, cuando el espacio de medidas tiene medida finita, la entropía de una transformación se puede calcular como el límite de la entropía de la transformación respecto a la sucesión de particiones.

Teorema 29. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$, $T : X \to X$ una transformación medible que preserva la medida μ. Si $(\alpha_n)_{n\in\mathbb{N}}$ es una sucesión de particiones medibles con,*

$$\bigvee_{n=1}^{\infty} S(\alpha_n) = S,$$

tales que α_{n+1} es un refinamiento de α_n para todo $n \in \mathbb{N}$, entonces,

$$h_\mu(T) = \lim_{n\to\infty} h_\mu(T, \alpha_n) = \sup_{n\in\mathbb{N}} h_\mu(T, \alpha_n).$$

Demostración. Por hipótesis, se tiene que α_{n+1} es un refinamiento de α_n para todo $n \in \mathbb{N}$, es decir, $\alpha_n \leq \alpha_{n+1}$. Esto implica que,

$$\bigvee_{i=0}^{n-1} T^{-i}(\alpha_m) \leq \bigvee_{i=0}^{n} T^{-i}(\alpha_m) \qquad \forall n, m \in \mathbb{N}.$$

Por la Proposición 31 se sigue que,

$$H_\mu\left(\bigvee_{i=0}^{n-1} T^{-i}(\alpha_m)\right) \leq H_\mu\left(\bigvee_{i=0}^{n} T^{-i}(\alpha_m)\right) \qquad \forall n, m \in \mathbb{N}.$$

Esto implica a su vez,

$$h_\mu(T, \alpha_m) \leq h_\mu(T, \alpha_{m+1}) \qquad \forall n, m \in \mathbb{N}.$$

Por el Lema 17 se sigue que,

$$h_\mu(T, \alpha_n) \leq h_\mu(T, \alpha_{n+1}) + H_\mu(\alpha_n|\alpha_{n+1}).$$

Así también, por el Lema 18,

$$H_\mu(\alpha_n|\alpha_{n+1}) \to 0 \qquad \text{Cuando } n \to \infty.$$

Entonces se tiene,

$$h_\mu(T, \alpha_n) \leq \lim_{n\to\infty} h_\mu(T, \alpha_{n+1}).$$

Como la sucesión es estrictamente creciente,

$$\lim_{n\to\infty} h_\mu(T, \alpha_{n+1}) = \sup_{n\in\mathbb{N}} h_\mu(T, \alpha_{n+1}).$$

También, por definición se tiene que,

$$\sup_{n\in\mathbb{N}} h_\mu(T, \alpha_{n+1}) \leq h_\mu(T).$$

Por lo tanto,

$$h_\mu(T) \leq \lim_{n\to\infty} h_\mu(T, \alpha_{n+1}) = \sup_{n\in\mathbb{N}} h_\mu(T, \alpha_{n+1}) \leq h_\mu(T).$$

□

4.4. Generadores

Calcular la entropía de una transformación medible a partir de la definición no siempre es posible. Sin embargo, en ocasiones es posible calcular la entropía de una transformación que preserva la medida, únicamente fijándonos en una sola partición medible. En esta sección, veremos que la entropía de una transformación coincide con la entropía de una transformación respecto a una partición medible, siempre y cuando la partición sea una *partición generadora* de la σ-álgebra. A continuación vamos a definir lo que entenderemos por un generador.

Definición 29. *Sea (X, S, μ) un espacio de medida. Sea $T : X \to X$ una transformación medible y ξ una partición medible de X.*

- *Decimos que ξ es un* ***generador lateral con respecto a*** *T si*

$$\bigvee_{k=0}^{+\infty} S(T^{-k}(\xi)) = S.$$

- *Decimos que ξ es un* ***generador bilateral con respecto a*** *T si*

$$\bigvee_{k=-\infty}^{+\infty} S(T^{-k}(\xi)) = S.$$

La mayoría de las veces es muy difícil encontrar generadores laterales y generadores bilaterales. Sin embargo, una vez que se encuentran, calcular la entropía de una transformación medible se vuelve mucho más sencillo. El Teorema de Kolmogorov-Sinaí afirma que la entropía de una transformación preservadora coincide con la entropía de la transformación con respecto a la partición generadora.

Teorema 30 (Kolmogorov-Sinai)**.** *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Si $T : X \to X$ es una transformación medible que preserva la medida μ, entonces se tienen las siguientes propiedades,*

- *Si ξ es un generador lateral, entonces $h_\mu(T) = h_\mu(T, \xi)$.*
- *Si ξ es un generador bilateral y T es invertible c.d(relμ), entonces $h_\mu(T) = h_\mu(T, \xi)$.*

Demostración. Supongamos primero que ξ es un generador lateral. Veamos que $h_\mu(T, \eta) \leq h_\mu(T, \xi)$ para toda partición medible η de X. Sea η una partición medible de X y consideramos la partición medible ξ_{n+1}. Por el Lema 17 se tiene que,

$$h_\mu(T, \eta) \leq h_\mu(T, \xi_{n+1}) + H_\mu(\eta|\xi_{n+1}). \tag{4.8}$$

Sin embargo, por definición tenemos,

$$\begin{aligned} h_\mu(T,\xi_{n+1}) &= \lim_{m\to\infty} \frac{1}{m} H_\mu\left(\bigvee_{k=0}^{m-1} T^{-k}(\xi_{n+1})\right) \\ &= \lim_{m\to\infty} \frac{1}{m} H_\mu(\xi_{m+n}). \end{aligned}$$

Por la Proposición 36 se tiene,

$$H_\mu(\xi_{m+n}) \leq H_\mu(\xi_m) + H_\mu(\xi_n).$$

Esto implica,

$$\begin{aligned} \lim_{m\to\infty} \frac{1}{m} H_\mu(\xi_{m+n}) &= \lim_{m\to\infty} \frac{1}{m} H_\mu(\xi_m) + \lim_{m\to\infty} \frac{1}{m} H_\mu(\xi_n) \\ &= \lim_{m\to\infty} \frac{1}{m} H_\mu(\xi_m). \end{aligned}$$

Así pues, se tiene que,

$$h_\mu(T,\xi_{n+1}) \leq h_\mu(T,\xi).$$

Por otro lado, como ξ_{n+1} es un refinamiento de ξ, por la Proposición 31,

$$H_\mu(\xi) \leq H_\mu(\xi_{n+1}).$$

Luego pues,

$$h_\mu(T,\xi) \leq h_\mu(T,\xi_{n+1}).$$

Entonces,

$$h_\mu(T,\xi_{n+1}) = h_\mu(T,\xi).$$

Luego, podemos escribir la desigualdad (4.8) como,

$$h_\mu(T,\eta) \leq h_\mu(T,\xi) + H_\mu(\eta|\xi_{n+1}). \tag{4.9}$$

Por hipótesis, ξ es un generador lateral y como para indices finitos sabemos que se cumple que,

$$S\left(\bigvee_{i=0}^{n} T^{-i}(\xi)\right) = \bigvee_{i=0}^{n} S(T^{-i}(\xi)).$$

Por el Lema 18,

$$H_\mu(\eta|\xi_{n+1}) \to 0 \qquad \text{si } n\to\infty.$$

Por lo tanto, por lo anterior y tomando límites en (4.9) se tiene que,

$$h_\mu(T,\eta) \leq h_\mu(T,\xi),$$

para toda partición medible η de X. Así pues, tomando supremos tenemos,

$$h_\mu(\eta) \leq h_\mu(T,\xi) \leq h_\mu(\eta),$$

es decir, $h_\mu(\eta) = h_\mu(T,\xi)$. Ahora para probar la segunda afirmación del teorema, supongamos que ξ es un generador bilateral y T es una transformación invertible $c.d.(rel.\mu)$. Al igual que con la primera afirmación, probaremos que $h_\mu(T,\eta) \leq h_\mu(T,\xi)$ para toda partición medible η de X.

Por hipótesis se tiene que T es invertible $c.d.(rel.\mu)$. Esto implica que $T^k(\xi)$ es una partición medible de X para toda $k \in \mathbb{N}$. Esto implica que,

$$\bigvee_{i=-n}^{n} T^{-i}(\xi),$$

es una partición medible de X. Por el Lema 17 se tiene que,

$$h_\mu(T,\eta) \leq h_\mu\left(T, \bigvee_{i=-n}^{n} T^{-i}(\xi)\right) + H_\mu\left(\eta \middle| \bigvee_{i=-n}^{n} T^{-i}(\xi)\right).$$

Sin embargo, como,

$$\begin{aligned}
\bigvee_{i=-n}^{n} T^{-i}(\xi) &= \bigvee_{l=0}^{2n} T^{-l+n}(\xi) \\
&= \bigvee_{l=0}^{2n} (T^n)^{-l}(\xi) \\
&= T^n \bigvee_{l=0}^{2n} T^{-l}(\xi) \\
&= T^n(\xi_{2n+1}).
\end{aligned}$$

Esto implica que,

$$\begin{aligned}
h_\mu\left(T, \bigvee_{i=-n}^{n} T^{-i}(\xi)\right) &= h_\mu\left(T, T^n \bigvee_{l=0}^{2n} T^{-l}(\xi)\right) \\
&= h_\mu(T, T^n(\xi_{2n+1})).
\end{aligned}$$

Como por hipótesis T preserva la medida μ,

$$h_\mu(T, T^n(\xi_{2n+1})) = h_\mu(T, \xi_{2n+1}).$$

Luego pues,

$$h_\mu(T,\eta) \leq h_\mu(T,\xi_{2n+1}) + H_\mu\left(\eta \middle| \bigvee_{i=-n}^{n} T^{-i}(\xi)\right).$$

Por un argumento análogo al que se dio en la prueba de la primera afirmación del teorema, se tiene que,

$$h_\mu(T, \xi_{en+1}) = h_\mu(T,\xi).$$

Así pues,

$$h_\mu(T,\eta) \leq h_\mu(T,\xi) + H_\mu\left(\eta \,\middle|\, \bigvee_{i=-n}^{n} T^{-i}(\xi)\right).$$

Por hipótesis ξ es bilateral, es decir,

$$\bigvee_{k=-\infty}^{+\infty} S(T^{-k}(\xi)) = S.$$

Por el Lema 18,

$$H_\mu\left(\eta \,\middle|\, \bigvee_{i=-n}^{n} T^{-i}(\xi)\right) \to 0 \qquad \text{si } n \to \infty.$$

Entonces,

$$h_\mu(T,\eta) \leq h_\mu(T,\xi),$$

para toda partición η de X. Finalmente, tomando supremos podemos concluir,

$$h_\mu(T) = h_\mu(T,\xi).$$

□

4.4.1. Cadenas de Markov

En esta sección vamos a calcular, con ayuda del Teorema de Kolmogorov-Sinaí, la entropía de las transformaciones corrimiento con respecto a la medida de Bernoulli y con respecto a la medida de Markov.

Medidas de Bernoulli

Sea $\sigma : \Sigma_k \to \Sigma_k$ la transformación corrimiento y μ la medida de Bernoulli bilateral en Σ_k que es generada por $p_1, \cdots, p_k$. Consideremos la siguiente partición de cilindros,

$$\xi = \{C_1, \cdots, C_k\}.$$

Observamos que,

$$\bigvee_{j=-n}^{n} \sigma^{-j}\xi = \{C_{i_{-n}\cdots i_n} : i_{-n}, \cdots, i_n \in \{i, \cdots k\}\}.$$

Como los cilindros generan a la σ-álgebra de Σ_k, por el Teorema de Kolmogorov-Sinaí, se tiene que,

$$h_\mu(\sigma) = h_\mu(\sigma,\xi).$$

Así pues,

$$
\begin{aligned}
h_\mu(\sigma) &= h_\mu(\sigma, \xi) \\
&= \lim_{n\to\infty} \frac{1}{n} H_\mu \left(\bigvee_{l=0}^{n-1} \sigma^{-l}(\xi) \right) \\
&= \lim_{n\to\infty} -\frac{1}{n} \sum_{i_0\cdots i_{n-1}} \mu(C_{i_0\cdots i_{n-1}}) \log \mu(C_{i_0\cdots i_{n-1}})) \\
&= \lim_{n\to\infty} -\frac{1}{n} \sum_{i_0\cdots i_{n-1}} p_{i_0} \cdots p_{i_{n-1}} \log(p_{i_0} \cdots p_{i_{n-1}}) \\
&= \lim_{n\to\infty} -\frac{1}{n} n \sum_{i=1}^{k} p_i \log p_i \\
&= \sum_{i=1}^{k} \psi(p_i).
\end{aligned}
$$

Como ψ es una función convexa, podemos utilizar la desigualdad de Jensen. Entonces,

$$
\begin{aligned}
-\sum_{i=1}^{k} \psi(p_i) &= -k - \sum_{i=1}^{k} \frac{1}{k} \psi(p_i) \\
&\le -k\psi\left(\sum_{i=1}^{k} \frac{p_i}{k} \right) \\
&= -k\psi\left(\frac{1}{k}\right) \\
&= -k\frac{1}{k} \log \frac{1}{k} \\
&= \log k.
\end{aligned}
$$

Por lo tanto, $h_\mu(\sigma) \le \log k$, con igualdad si y sólo si

$$p_i = \cdots = p_k = \frac{1}{k}.$$

Consideremos la transformación $\sigma_1 : \{1,2,3\}^{\mathbb{Z}} \to \{1,2,3\}^{\mathbb{Z}}$ con la medida de Bernoulli con probabilidades $\frac{1}{3}, \frac{1}{3}, \frac{1}{3}$. También consideremos la transformación $\sigma_1 : \{1,2\}^{\mathbb{Z}} \to \{1,2\}^{\mathbb{Z}}$ con la medida de Bernoulli con probabilidades $\frac{1}{2}, \frac{1}{2}$. Por una parte,

$$h_\mu(\sigma_1) = \log 3.$$

Por otra parte,

$$h_\mu(\sigma_2) = \log 2.$$

Por lo tanto, σ_1 y σ_2 tienen entropía diferente. Sin embargo, consideremos la transformación $\sigma_1 : \{1,2,3,4\}^{\mathbb{Z}} \to \{1,2,3,4\}^{\mathbb{Z}}$ con la medida de Bernoulli con probabilidades $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$. Así también, consideremos la transformación $\sigma_2 : \{1,2,3,4,5\}^{\mathbb{Z}} \to \{1,2,3,4,5\}^{\mathbb{Z}}$ con la medida de Bernoulli con probabilidades $\frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}$,. Por una parte,

$$h_\mu(\sigma_1) = \log 4 = \log 2^2 = 2\log 2.$$

Por otra parte,

$$\begin{aligned}
h_\mu(\sigma_2) &= -\sum_{i=1}^{k} p_i \log p_i \\
&= -\frac{1}{2} - 4\frac{1}{8}\log\frac{1}{8} \\
&= \frac{1}{2}\log 2 + \frac{1}{2}\log 2^3 \\
&= \frac{1}{2}(\log 2^4) \\
&= 2\log 2.
\end{aligned}$$

Por lo tanto, σ_1 y σ_2 tienen la misma entropía.

Medidas de Markov

Sea P una matriz estocástica irreducible y p el vector de probabilidad asociado a P tal que (P,p) es un par estocástico. Sea $\sigma : \Sigma_k \to \Sigma_k$ la transformación corrimiento y μ la medida de Markov bilateral asociada al par estocástico (P,p). Consideremos la misma partición de cilindros que utilizamos para la medida de Bernoulli.

$$\xi = \{C_1, \cdots, C_k\}.$$

Dado que la partición es una partición generadora, vamos a utilizar el Teorema de Kolmogorov-Sinaí.

$$\begin{aligned}
H_\mu\left(\bigvee_{l=0}^{n-1} \sigma^{-l}\xi\right) &= -\sum_{i_0\cdots i_{n-1}} \mu(C_{i_0\cdots i_{n-1}})\log\mu(C_{i_0\cdots i_{n-1}}) \\
&= -\sum_{i_0\cdots i_{n-1}} p_{i_0}p_{i_0i_1}\cdots p_{i_{n-2}i_{n-1}}\log(p_{i_0}p_{i_0i_1}\cdots p_{i_{n-2}i_{n-1}}) \\
&= -\sum_{i_0\cdots i_{n-1}} p_{i_0}p_{i_0i_1}\cdots p_{i_{n-2}i_{n-1}}\log p_{i_0} \\
&- \sum_{i_0\cdots i_{n-1}} p_{i_0}p_{i_0i_1}\cdots p_{i_{n-2}i_{n-1}}\sum_{j=0}^{n-2}\log p_{i_ji_{j+1}} \\
&= -\sum_{i=1}^{k} p_i\log p_i - (n-1)\sum_{i=1}^{k}\sum_{j=1}^{k} p_ip_{ij}\log p_{ij}.
\end{aligned}$$

Por el Teorema de Kolmogorov-Sinaí,

$$\begin{aligned} h_\mu(\sigma) &= \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{l=0}^{n-1} \sigma^{-l}\xi\right) \\ &= -\sum_{i=1}^{k}\sum_{j=1}^{k} p_i p_{ij} \log p_{ij}. \end{aligned}$$

Sustituimos $\psi(p_{ij}) = p_{ij}\log p_{ij}$. Como ψ es una función convexa, podemos usar la desigualdad de Jensen,

$$\begin{aligned} h_\mu(\sigma) &= -\sum_{i=1}^{k}\sum_{j=1}^{k} p_i \psi(p_{ij}) \\ &= -\sum_{i=1}^{k} p_i \left(\sum_{j=1}^{k} \psi(p_i j)\right) \\ &\leq -\sum_{i=1}^{k} p_i k \psi\left(\sum_{j=1}^{k} \frac{p_{ij}}{k}\right) \\ &= -\sum_{i=1}^{k} k\psi\left(\frac{1}{k}\right) \\ &= \log k. \end{aligned}$$

Por lo tanto, $h_\mu(\sigma) \leq \log k$ con igualdad si y sólo si $p_{ij} = \frac{1}{k}$ para cada i, j.

4.5. El Teorema de Shannon-McMillan-Breiman

Los resultados que hemos desarrollado en las secciones anteriores nos permitirán ahora formular el resultado principal de este capítulo, que es el Teorema de Shannon-McMillan-Breiman. A continuación vamos a desarrollar algunas definiciones ulteriores y pruebas auxiliares para facilitar la enunciación y prueba del Teorema de Shannon-McMillan-Breiman.

Consideremos una partición medible ξ del espacio X. Como ξ es una partición, por definición sabemos que para casi todo $x \in X$, existe un único elemento de la partición $C \in \xi$ tal que $x \in C$. Para hacer referencia al elemento C al cual pertenece x, lo denotaremos como $\xi(x)$.

Observemos que, dado un espacio medible X, si ξ es una partición medible de X y consideramos la σ-álgebra generada por la partición ξ, es decir, $S(\xi)$ podemos dar una caracterización de la esperanza condicional de una función integrable. Sea $f \in L^1(X,\mu)$, la esperanza condicional de f es la siguiente,

$$E(f|S(\xi)) = \sum_{B\in\beta} \chi_B \frac{1}{\mu(B)} \int_B f \mathrm{d}\mu.$$

En efecto, para verificar que la expresión anterior es la esperanza condicional de la función f, basta integrar. Al integrar la expresión anterior vemos que se satisface la propiedad fundamental de la esperanza condicional, es decir,

$$\int E(f|S(\xi))\mathrm{d}\mu = \int f\mathrm{d}\mu.$$

En particular, si $A \subseteq X$ es un conjunto medible, su función característica χ_A es integrable. Por la observación anterior, la esperanza condicional de χ_A con respecto a la σ-subálgebra $S(\xi)$ es la siguiente,

$$\begin{aligned} E(\chi_A|S(\xi)) &= \sum_{B\in\beta} \chi_B \frac{1}{\mu(B)} \int_B \chi_A \mathrm{d}\mu \\ &= \sum_{B\in\beta} \chi_B \frac{1}{\mu(B)} \mu(A \cap B) \\ &= \mu(A|\xi). \end{aligned}$$

Definimos la siguiente sucesión de particiones medibles,

$$\begin{aligned} \alpha_1 &= X \\ \alpha_n &= \bigvee_{k=1}^{n-1} T^{-k}(\xi). \end{aligned}$$

Lema 19. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Sea $T : X \to X$ una transformación que preserva la medida μ y ξ una partición medible de X. Definimos para cada $n \in \mathbb{N}$,*

$$F_n = -\sum_{A\in\xi} \chi_A \log \mu(A|\alpha_n).$$

Así también, definimos,

$$F^* = \sup_{n\geq 1} |F_n|.$$

Entonces, para cada $t \geq 0$ y para cada $A \in \xi$ se tiene lo siguiente:

- $\mu(\{x \in A : F^*(x) > t\}) \leq e^{-t}$.
- $F^* \in L^1(X, \mu)$.

Demostración. Sea $t \geq 0$ y $A \in \xi$. Para cada $n \geq 1$ definimos,

$$F_n^A(x) = -\log \mu(A|\alpha_n)(x).$$

Así también,

$$B_n^A = \{x \in X : F_k^A(x) \leq t \ \forall k \in \{1, 2, \cdots, n-1\} \text{ y } F_n^A(x) > t\}.$$

Observamos que si $x \in B_n^A$, entonces,

$$\begin{aligned} F_n^A(x) &> t \\ -\log \mu(A|\alpha_n) &> t \\ \log \mu(A|\alpha_n) &< -t \\ \mu(A|\alpha_n) &< e^{-t}. \end{aligned}$$

Como $B_n^A \in S(\alpha_n)$, por la propiedad fundamental de la esperanza condicional,

$$\begin{aligned} \mu(B_n^A \cap A) &= \int_{B_n^A} \chi_A \mathrm{d}\mu \\ &= \int_{B_n^A} \mu(A|\alpha_n) \mathrm{d}\mu \\ &\leq \int_{B_n^A} e^{-t} \mathrm{d}\mu \\ &= e^{-t}\mu(B_n^A). \end{aligned}$$

Entonces,

$$\begin{aligned} \mu(\{x \in A : F^*(x) > t\}) &= \mu(\{x \in A : F_n(x) > t, \text{ para algún } n\}) \\ &= \mu(\{x \in A : F_n^A(x) > t, \text{ para algún } n\}) \\ &= \mu\left(\bigcup_{n=1}^{\infty} (A \cap B_n^A)\right) \\ &= \sum_{n=1}^{\infty} \mu(A \cap B_n^A) \\ &\leq e^{-t} \sum_{n=1}^{\infty} \mu(B_n^A) \\ &\leq e^{-t}. \end{aligned}$$

Ahora, veamos que $F^* \in L^1(X, \mu)$. Notamos que,

$$\mu(\{x \in A : F^*(x) > t\}) \leq \mu(A).$$

Esto implica que,

$$\mu(\{x \in A : F^*(x) > t\}) \leq \text{mín}(\mu(A), e^{-t}).$$

Luego pues,

$$\begin{aligned}\int F^*(x)\mathrm{d}\mu(x) &= \int_0^\infty \mu(\{x\in X : F^*(x) > t\})\mathrm{d}t\\ &= \int_0^\infty \sum_{A\in\xi}\mu(\{x\in A : F^*(x) > t\})\mathrm{d}t\\ &= \sum_{A\in\xi}\int_0^\infty \mu(\{x\in A : F^*(x) > t\})\mathrm{d}t\\ &\le \sum_{A\in\xi}\int_0^\infty \text{mín}(\mu(A), e^{-t})\mathrm{d}t\\ &= \sum_{A\in\xi}\int_0^{-\log\mu(A)} \mu(A)\mathrm{d}t + \sum_{A\in\xi}\int_{-\log\mu(A)}^\infty e^{-t}\mathrm{d}t\\ &= -\sum_{A\in\xi}\log\mu(A) + \sum_{A\in\alpha}\frac{\mu(A)}{\log e}\\ &= H_\mu(\xi) + 1 < \infty.\end{aligned}$$

Por lo tanto, $F^* \in L^1(X,\mu)$. □

A continuación veremos un lema auxiliar que se utilizará en la prueba del Teorema de Shannon-McMillan-Breiman.

Lema 20. *Sea (X,S,μ) un espacio de medida finita, sea $T : X \to X$ una transformación medible y ξ una partición medible de X, entonces se tiene la siguiente identidad,*

$$\log\mu(\xi_n(x)) = \log\mu(\xi(T^{n-1}(x))) + \sum_{j=1}^{n-1}\log\mu\left(\xi(T^{j-1}(x))\,\middle|\,\bigvee_{k=1}^{n-j}T^{-k}\xi\right).$$

Demostración. Por inducción sobre n. Observamos que la base de la inducción con $n = 1$ se satisface de manera evidente. Por hipótesis de inducción, supongamos que la identidad se satisface para algún n natural y veamos que se satisface para $n+1$. Recordemos que se había definido α_n como,

$$\alpha_n = \bigvee_{k=1}^{n-1}T^{-k}\xi.$$

Asimismo, si $A \subseteq X$ es un conjunto medible, la esperanza condicional de χ_A con respecto a $S(\xi)$ se denota como $\mu(A\vee\xi)$ y se caracteriza como,

$$\mu(A\vee\xi) = \sum_{B\in\alpha_n}\frac{\mu(A\cap B)}{\mu(B)}\chi_B \qquad \forall A\in\xi \qquad c.d.rel(\mu).$$

Esto implica que,

$$\log \mu(A \vee \xi) = \sum_{B\in\alpha_n} \log \frac{\mu(A\cap B)}{\mu(B)}\chi_B \qquad \forall A\in\xi \qquad c.d.rel(\mu).$$

Así pues, por definición, por propiedades del logaritmo y por la caracterización de la esperanza condicional de χ_A se tiene lo siguiente,

$$\begin{aligned}
\log \mu(\xi_{n+1}(x)) &= \sum_{A\in\xi}\sum_{B\in\alpha_n} \log \mu(A\cap B)\chi_{A\cap B}(x)\\
&= \sum_{A\in\xi}\sum_{B\in\alpha_n} \log \mu(B)\frac{\mu(A\cap B)}{\mu(B)}\chi_A(x)\chi_B(x)\\
&= \sum_{A\in\xi}\sum_{B\in\alpha_n} \log \mu(B)\chi_A(x)\chi_B(x)\\
&+ \sum_{A\in\xi}\sum_{B\in\alpha_n} \frac{\mu(A\cap B)}{\mu(B)}\chi_A(x)\chi_B(x)\\
&= \sum_{A\in\xi}\sum_{B\in\alpha_n} \log \mu(B)\chi_A(x)\chi_B(x) + \sum_{A\in\xi}\sum_{B\in\alpha_n} \mu(A\vee\alpha_n)\chi_A(x)\\
&= \sum_{B\in\alpha_n} \log \mu(B)\chi_B(x) + \sum_{A\in\xi}\sum_{B\in\alpha_n} \mu(A\vee\alpha_n)\chi_A(x)\\
&= \log\mu(\alpha_n(x)) + \log\mu(\xi(x)\vee\alpha_n).
\end{aligned}$$

Como $\alpha_n(x) = \xi_n(T(x))$, entonces se tiene que,

$$\log \mu(\xi_{n+1}(x)) = \log\mu(\xi_n(T(x))) + \log\mu(\xi(x)\vee\alpha_n).$$

Por la hipótesis de inducción se obtiene lo siguiente,

$$\begin{aligned}
\log \mu(\xi_{n+1}(x)) &= \log\mu(\xi(T^{n-1}(T(x)))) + \sum_{j=1}^{n-1}\log\mu(\xi(T^{j-1}(T(x)))|\alpha_{n+1-j})\\
&+ \log\mu(\xi(x)\vee\alpha_n)\\
&= \log\mu(\xi(T^n(x))) + \sum_{j=2}^{n-1}\log\mu(\xi(T^{j-1}(x))|\alpha_{n+1-j})\\
&+ \log\mu(\xi(x)\vee\alpha_n)\\
&= \log\mu(\xi(T^n(x))) + \sum_{j=1}^{n-1}\log\mu(\xi(T^{j-1}(x))|\alpha_{n+1-j}).
\end{aligned}$$

□

El siguiente lema afirma que si ϕ es una función integrable, entonces $\frac{\phi\circ T^n}{n}$ converge a 0 para casi todos los puntos de X.

Lema 21. *Sea (X, S, μ) un espacio de medida y $T : X \to X$ una transformación que preserva la medida μ en X. Si $\phi \in L^1(X, \mu)$, entonces,*

$$\lim_{n\to\infty} \frac{\phi(T^n(x))}{n} = 0 \qquad c.d.\ (rel\mu).$$

Demostración. Sea $A \subseteq X$ un conjunto medible. Consideramos la función $\chi_A \in L^1(X, \mu)$, como,

$$\chi_A(T^n(x)) \leq 1 \qquad \forall x \in X \text{ y } \forall n \in \mathbb{N}.$$

Entonces,

$$\frac{\chi_A(T^n(x))}{n} \leq \frac{1}{n} \to 0 \qquad \forall x \in X.$$

Es decir, la propiedad se cumple para las funciones características. Consideremos una función simple s. Sabemos que las funciones simples tienen la forma,

$$s = \sum_{i=1}^{k} c_i \chi_{E_i},$$

donde, $E_i \in S$ y $c_i \in \mathbb{R}$ para cada i. Como,

$$\begin{aligned}
\lim_{n\to\infty} \frac{s(T^n(x))}{n} &= \lim_{n\to\infty} \sum_{i=1}^{k} \frac{c_i \chi_{E_i}(T^n(x))}{n} \\
&= \sum_{i=1}^{k} c_i \lim_{n\to\infty} \frac{\chi_{E_i}(T^n(x))}{n} \\
&= 0 \qquad \forall x \in X.
\end{aligned}$$

Consideremos $\phi \in L^1(X, \mu)$. Sabemos que $\phi = \phi^+ - \phi^-$, donde ϕ^+, ϕ^- son funciones medibles. Como ϕ es integrable, entonces sabemos que es acotada c.d. (rel μ) y entonces ϕ^+, ϕ^- también son funciones acotadas c.d. (rel μ). Por el Lema Básico de Aproximación, sabemos que existe una sucesión de funciones simples $(s_l)_{l\in\mathbb{N}}$ tales que,

$$\lim_{l\to\infty} s_l = \phi^+.$$

Más aun, sabemos que la convergencia es uniforme.

$$\begin{aligned}
\lim_{n\to\infty} \frac{\phi^+(T^n(x))}{n} &= \lim_{n\to\infty} \lim_{l\to\infty} \frac{1}{n} s_l((T^n(x))) \\
&= \lim_{l\to\infty} \lim_{n\to\infty} \frac{1}{n} s_l((T^n(x))) \\
&= 0 \qquad c.d. \text{ rel}(\mu).
\end{aligned}$$

Procediendo de manera análoga, se prueba,

$$\lim_{n\to\infty} \frac{\phi^-(T^n(x))}{n} = 0 \qquad c.d. \text{ rel}(\mu).$$

Por lo tanto,

$$\lim_{n\to\infty} \frac{\phi(T^n(x))}{n} = 0 \qquad c.d. \text{ rel}(\mu).$$

$\square$

El siguiente teorema que vamos a probar es el Teorema de Shannon-McMillan-Breiman. El teorema muestra una manera de calcular la entropía de una transformación medible. Para probar el Teorema de Shannon-McMillan-Breiman vamos a ocupar dos resultados importantes de la Teoría de la Medida, a saber, el Teorema de la Convergencia Dominada y el Teorema del Incremento Martingale. El primero ya fue enunciado anteriormente como el Teorema 17 A continuación enunciamos sin prueba el segundo.

Teorema 31 (Incremento Martingale)**.** *Si $\phi : X \to \mathrm{R}$ es una función integrable y F_n es una sucesión de σ-álgebras no decreciente tales que $\bigcup_{n=1}^{\infty} F_n$ genera la σ-álgebra F, entonces la sucesión de funciones (ϕ_{F_n}) converge a F casi dondequiera y en L^1.*

- $\phi_{F_n} \to \phi_F$ *c.d. rel(μ).*
- $$\int |\phi_{F_n} - \phi_F| \, \mathrm{d}\mu \to 0 \qquad si\ n \to \infty.$$

Para probar el Teorema de Shannon-McMillan-Breiman ocuparemos el Teorema Ergódico de Birkhoff. Recordamos que habíamos visto que el Teorema Ergódico se podía fortalecer al añadir la hipótesis de que la transformación medible era una transformación ergódica. La versión que enunciamos a continuación del Teorema de Shannon-McMillan-Breiman no supone la ergodicidad de la transformación. Posteriormente hablaremos de la versión de Teorema que supone que la transformación es ergódica. La segunda versión, aunque menos general, será la que ocuparemos en el siguiente capítulo cuando hablemos del problema de la compresión.

Teorema 32 (Shannon-McMillan-Breiman)**.** *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$. Sea $T : X \to X$ una transformación que preserva la medida μ y ξ una partición medible de X. Entonces el límite*

$$h_\mu(T, \xi, x) := \lim_{n\to\infty} -\frac{1}{n} \log \mu(\xi_n(x))$$

existe c.d. rel.(μ). Más aún, la función

$$x \mapsto h_\mu(T, \xi, x)$$

tiene las siguientes propiedades:

- *es T-invariante c.d.(rel.μ).*
- *es μ integrable.*

- $$h_\mu(T,\xi) = \int h_\mu(T,\xi,x)\mathrm{d}\mu(x).$$

Demostración. Sea $A \subseteq X$ un conjunto medible. Consideramos la función característica $\chi_A \in L^1(X,\mu)$ y la σ-subálgebra $S(\xi)$ generada por la partición medible ξ de X. Por la Proposición 14, tenemos que existe la esperanza condicional χ_A con respecto a la σ-subálgebra $\mathcal{B}(\xi)$ generada por ξ y la denotamos como $\mu(A|\xi)$. Por el Lema 20 se tiene,

$$\log\mu(\xi_n(x)) = \log\mu(\xi(T^{n-1}(x))) + \sum_{j=1}^{n-1}\log\mu\left(\xi(T^{j-1}(x))\middle|\bigvee_{k=1}^{n-j}T^{-k}\xi\right)$$

Recordamos que se definió la siguiente sucesión de particiones medibles,

$$\begin{aligned}\alpha_1 &= X\\ \alpha_n &= \bigvee_{k=1}^{n-1}T^{-k}(\xi) \qquad \forall n\in\mathbb{N}.\end{aligned}$$

Finalmente, definimos $F_k : X \to \mathbb{R}$ como sigue,

$$F_k(x) = \log\mu(\xi(x)|\alpha_k).$$

Sustituyendo obtenemos,

$$\begin{aligned}\log\mu(\xi_n(x)) &= \log\mu(\xi(T^{n-1}(x))) + \sum_{j=1}^{n-1}\log\mu\left(\xi(T^{j-1}(x))\middle|\bigvee_{k=1}^{n-j}T^{-k}\xi\right)\\ &= \log\mu(\xi(T^{n-1}(x))|\alpha_1) + \sum_{j=1}^{n-1}\log\mu(\xi(T^{j-1}(x))|\alpha_{n-j+1})\\ &= F_1(T^{n-1}(x)) + \sum_{j=1}^{n-1}F_{n-j}(T^j(x))\\ &= \sum_{j=0}^{n-1}F_{n-j}(T^j(x)).\end{aligned}$$

Así pues,

$$\log\mu(\xi_n(x)) = \sum_{j=0}^{n-1}F_{n-j}(T^j(x)).$$

Consideremos la función $\phi : X \to \mathbb{R}$ definida como sigue:

$$\phi(x) = \log\mu(\xi(x)) \qquad c.d.\ \mathrm{rel}(\mu).$$

Observamos que $\phi \in L^1(X,\mu)$, pues,

$$\int|\phi|\mathrm{d}\mu = -\sum_{C\in\xi}\mu(C)\log\mu(C) = H_\mu(\xi) < +\infty.$$

Definimos,

$$\alpha_\infty = \bigvee_{n=1}^{\infty} S(\alpha_n).$$

Por el Teorema del Incremento Martingale se sigue,

$$\mu(\xi(x)|\alpha_n) \xrightarrow{L^1} \mu(\xi|\alpha_\infty) \qquad c.d.\ \mathrm{rel}(\mu).$$

Esto implica que,

$$\log \mu(\xi(x)|\alpha_n) \xrightarrow{L^1} \log \mu(\xi|\alpha_\infty) \qquad c.d.\mathrm{rel}(\mu).$$

Entonces,

$$F_k \xrightarrow{L^1} F \qquad c.d.(rel.\mu).$$

Donde $F(x) = \log \mu(\xi(x)|\alpha_\infty)$. Como,

$$\begin{aligned}\frac{1}{n}\log \mu(\xi_n(x)) &= \frac{1}{n}\sum_{j=0}^{n-1} F_{n-j}(T^j(x)) \\ &= \frac{1}{n}\sum_{j=0}^{n-1} F(T^j(x)) + \frac{1}{n}\sum_{j=0}^{n-1}(F_{n-j} - F)(T^j(x)).\end{aligned}$$

Por el Teorema Ergódico de Birkhoff sabemos que existe $\psi : X \to X$ tal que,

$$\psi(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} F(T^j(x)).$$

Donde la función ψ satisface las siguientes propiedades,

- ψ es T-invariante c.d. rel(μ).
- $\psi \in L^1(X, \mu)$.
- $$\int \psi \mathrm{d}\mu = \int F \mathrm{d}\mu.$$

Veamos que,

$$\frac{1}{n}\sum_{j=0}^{n-1}(F_{n-j} - F)(T^j(x)) \to 0 \qquad \text{si } n \to \infty \ c.d.\ \mathrm{rel}(\mu).$$

Definimos,

$$F^* = \sup_{n\geq 1} |F_k|.$$

Por el Lema 19 se tiene que $F^* \in L^1(X,\mu)$. Para cada k natural, definimos $G_k : X \to X$ de la siguiente manera,

$$G_k = |F_k - F|.$$

También definimos $G_k^* : X \to X$ como,

$$G_k^* = \sup_{n\geq k} G_n.$$

Así pues,

$$\begin{aligned}\frac{1}{n}\sum_{j=0}^{n-1} G_{n-j}(T^j(x)) &= \frac{1}{n}\sum_{j=0}^{n-k} G_{n-j}(T^j(x)) + \frac{1}{n}\sum_{j=n-k+1}^{n-1} G_{n-j}(T^j(x)) \\ &\leq \frac{1}{n}\sum_{j=0}^{n-k} G_k^*(T^j(x)) + \frac{1}{n}\sum_{j=n-k+1}^{n-1} (F^*+F)(T^j(x))\end{aligned}$$

Por el Lema 19, por el Lema 21 y por el Teorema Ergódico de Birkhoff se tiene que,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} G_{n-j}(T^j(x)) &\leq \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-k} G_{n-j}(T^j(x)) \\ &= \psi_k(x) \qquad c.d.\ \mathrm{rel}(\mu),\end{aligned}$$

donde para cada $k \in \mathbb{N}$ se tiene lo siguiente,

- ψ_k es T-invariante c.d. rel(μ).
- $\psi_k \in L^1(X,\mu)$.
- $$\int \psi_k \mathrm{d}\mu = \int G_k^* \mathrm{d}\mu.$$

Como $G_k^* \to 0$ c.d. rel(μ) si $k \to \infty$ y además,

$$G_k^* \leq F * + F \qquad \forall k \in \mathbb{N}.$$

Con $F * + F \in L^1(X,\mu)$. Por el Teorema de la Convergencia Dominada,

$$\begin{aligned}\int \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}(G_{n-j}\circ T^j)\mathrm{d}\mu &\leq \lim_{k\to\infty}\int G_k^*\mathrm{d}\mu \\ &= \int 0 \mathrm{d}\mu \\ &= 0.\end{aligned}$$

Por lo tanto,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1}(G_{n-j}(T^j(x))) = 0 \qquad c.d.\ \mathrm{rel}(\mu).$$

Luego pues,

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log\mu(\xi_n(x)) &= \lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}F(T^j(x))\\ &= \psi(x) \qquad c.d.\ \mathrm{rel}(\mu).\end{aligned}$$

De esta manera, tenemos la primera parte del teorema, a saber, que el límite existe,

$$h_\mu(X,\xi,x) = \lim_{n\to\infty} -\frac{1}{n}\log\mu(\xi_n(x)) \qquad c.d.\ \mathrm{rel}(\mu).$$

Además, $h_\mu(X,\xi,\mu)$ es T-invariante y pertenece a $L^1(X,\mu)$. Más aún,

$$\begin{aligned}\int \lim_{n\to\infty} -\frac{1}{n}\log\mu(\xi_n(x))\mathrm{d}\mu(x) &= -\int\psi\mathrm{d}\mu\\ &= -\int F\mathrm{d}\mu.\end{aligned}$$

Como F_n converge a F en L^1 cuando $n\to\infty$ y por el Teorema 28, entonces,

$$\begin{aligned}-\int F\mathrm{d}\mu &= -\lim_{n\to\infty}\int F_n\mathrm{d}\mu\\ &= \lim_{n\to\infty}\int -\log\mu(\xi(x)|\alpha_n)\mathrm{d}\mu\\ &= \lim_{n\to\infty}\sum_{A\in\xi}\int_A -\log\mu(A|\alpha_n)\mathrm{d}\mu\\ &= \lim_{n\to\infty}\sum_{A\in\xi}\int_A\sum_{B\in\alpha_n} -\log\frac{\mu(A\cap B)}{\mu(B)}\chi_B\mathrm{d}\mu\\ &= \lim_{n\to\infty} -\sum_{A\in\xi}\sum_{B\in\alpha_n}\mu(A\cap B)\log\frac{\mu(A\cap B)}{\mu(B)}\\ &= \lim_{n\to\infty} H_\mu(\xi|\alpha_n)\\ &= h_\mu(T,\xi).\end{aligned}$$

□

El Teorema de Shannon-McMillan-Breiman se puede fortalecer en gran medida añadiendo algunas hipótesis adicionales. Si suponemos que la transformación T es una transformación ergódica, entonces utilizando el Teorema 21 en lugar del Teorema Ergódico de Birkhoff en la prueba del Teorema de Shannon-McMillan-Breiman, podemos concluir que,

$$h_\mu(T,\xi) = \lim_{n\to\infty} -\frac{1}{n}\log\mu(\xi_n(x)).$$

Asimismo, si añadimos la hipótesis de que la partición ξ es una partición generadora, por el Teorema de Kolomogorov-Sinaí, podemos concluir que,

$$h_\mu(T) = \lim_{n\to\infty} -\frac{1}{n}\log\mu(\xi_n(x)).$$

Ejemplo 16. Consideremos el alfabeto $\mathcal{A} = \{0,1\}$ y espacio de $\Sigma_2^+ = \mathcal{A}^{\mathbb{N}}$. Supongamos que $p_0 = p(0) = r$ y $p_1 = p(1) = s$ y consideremos la medida de Bernoulli. Entonces la probabilidad de que una secuencia inicie de la siguiente manera,

$$(101101).$$

es igual a,

$$\begin{aligned}\mu(C_{101101}) &= p_1p_0p_1p_1p_0p_1\\ &= r^2s^4.\end{aligned}$$

El Teorema de Shannon-McMillan-Breiman afirma que cualquier sucesión de longitud 2^n tiene aproximadamente la misma probabilidad. Tomando el logaritmo base 2,

$$\mu(C_{i_0i_1i_2i_3i_4i_5i_6}) \approx 2^{-nh},$$

donde h es la entropía de la transformación corrimiento σ.

Capítulo 5

Compresión de Datos

En este capítulo se probará, a partir del Teorema de Shannon-McMillan-Breiman lo que se conoce como la *Propiedad de la Equipartición Asintótica* (AEP). La AEP afirma que la mayoría de los conjuntos que pertenecen a una partición ξ_n tienen aproximadamente la misma medida. Posteriormente, con base en la AEP encontraremos una cota a la compresión en función de la entropía. Finalmente, estudiaremos el algoritmo de Huffman para alcanzar una compresión óptima de un texto.

5.1. Propiedad de la Equipartición Asintótica

A partir del Teorema de Shannon-McMillan-Breiman podemos dividir el espacio en dos clases de conjuntos, a saber, los *conjuntos típicos* y los *conjuntos no típicos*. Los conjuntos típicos son la mayoría y su medida es muy cercana al total, mientras que los conjuntos no típicos tienen la característica de que su medida es muy pequeña, de manera podemos ignorarlos. Dicha propiedad recibe el nombre de la Propiedad de la Equipartición Asintótica.

A continuación haremos algunas observaciones respecto del Teorema de Shannon-McMillan-Breiman y posteriormente probaremos la existencia de los conjuntos típicos y la de los conjuntos no típicos.

Consideremos un espacio de medida, una transformación medible T y una partición medible. El Teorema de Shannon-McMillan-Breiman afirma que en el caso de que el espacio sea de medida finita X, la transformación T sea ergódica y la partición medible generadora, se tiene que,

$$-\frac{1}{n}\log\mu(\xi(x)) \to h \qquad \textit{c.d.} \text{ (rel } \mu).$$

Ahora bien, el *Teorema de Egorov* afirma que en espacios de medida finita la convergencia c.d. rel (μ) implica la convergencia en medida. Así pues, el Teorema de Shannon-McMillan-Breiman implica,

$$-\frac{1}{n}\log\mu(\xi(x)) \xrightarrow{\mu} h.$$

A continuación, probaremos la AEP y con ello la existencia de los conjuntos típicos. Como convención vamos a suponer que D es la base del logaritmo.

Teorema 33 (Propiedad de la Equipartición Asintótica (AEP))**.** *Sea* (X, S, μ) *un espacio de medida tal que* $\mu(X) = 1$*,* $T : X \to X$ *es una transformación ergódica y* ξ *una partición medible generadora de* X*. Entonces para cada* $\epsilon > 0$ *y* $n \geq 1$ *existe un subconjunto de* ξ_n *que denotamos como* A_ϵ^n *tales que,*

$$D^{-n(h+\epsilon)} < \mu(A_\epsilon^n(x)) < D^{-n(h-\epsilon)}.$$

Los conjuntos $A_\epsilon^n(x)$ *reciben el nombre de* ***conjuntos típicos****.*

Demostración. Por el Teorema de Shannon-McMillan-Breiman se tiene que para cada $\epsilon > 0$ existe $N \in \mathbb{N}$ tal que $\forall n \geq N$ se tiene,

$$\mu\left(\left\{x : \left|-\frac{1}{n}\log\mu(\xi_n(x)) - h\right| \geq \epsilon\right\}\right) < \epsilon.$$

Así pues, podemos dividir el espacio X como,

$$X = \left\{x : \left|-\frac{1}{n}\log\mu(\xi_n(x)) - h\right| \geq \epsilon\right\} \cup \left\{x : \left|-\frac{1}{n}\log\mu(\xi_n(x)) - h\right| < \epsilon\right\}.$$

Como la función,

$$x \mapsto \mu(\xi(x)),$$

es constante para cada $x \in \xi_n(x)$, entonces se tiene que los elementos de la partición pertenecen a uno u a otro de los conjuntos. Denotamos a los elementos de la partición que pertenecen al conjunto,

$$\left\{x : \left|-\frac{1}{n}\log\mu(\xi_n(x)) - h\right| < \epsilon\right\},$$

como A_ϵ^n. Veamos que satisfacen el enunciado de la proposición. Por definición, un miembro x del conjunto satisface lo siguiente,

$$\left|-\frac{1}{n}\log\mu(A_\epsilon^n(x)) - h\right| < \epsilon.$$

Así pues,

$$-\epsilon + h < -\frac{1}{n}\log\mu(A_\epsilon^n(x)) < \epsilon + h$$

$$-n(h+\epsilon) < \log\mu(A_\epsilon^n(x)) < -n(h-\epsilon)$$

$$D^{-n(h+\epsilon)} < \mu(A_\epsilon^n(x)) < D^{-n(h-\epsilon)}.$$

□

Más aun, dado $\epsilon > 0$ y $n \geq 1$ la unión de los conjuntos típicos es el conjunto,

$$\bigcup_{B \in A_\epsilon^n} B = \left\{ x : \left| -\frac{1}{n} \log \mu(\xi_n(x)) - h \right| < \epsilon \right\}.$$

El Teorema de Shannon-McMillan-Breiman afirma que la medida del complemento de los conjuntos típicos es menor a ϵ, entonces se tiene que,

$$\mu \left(\left\{ x : \left| -\frac{1}{n} \log \mu(\xi_n(x)) - h \right| < \epsilon \right\} \right) > 1 - \epsilon.$$

A continuación se probará una importante propiedad de los conjuntos típicos que se basa en esta observación. La siguiente proposición nos da información acerca de la cardinalidad de los conjuntos típicos.

Proposición 40. *Sea (X, S, μ) un espacio de medida tal que $\mu(X) = 1$, $T : X \to X$ es una transformación ergódica y ξ una partición medible generadora de X. Entonces,*

$$(1 - \epsilon) D^{n(h-\epsilon)} < |A_\epsilon^n| < D^{n(h+\epsilon)}.$$

Demostración. Sea ϵ y $n \geq 1$. Dado que los elementos de A_ϵ^n son subconjuntos de la partición ξ_n, entonces,

$$\mu \left(\bigcup_{B \in A_\epsilon^n} B \right) = \sum_{B \in A_\epsilon^n} \mu(B).$$

Así pues, por la Propiedad de la Equipartición Asintótica se tiene lo siguiente,

$$\begin{aligned} 1 > \mu & \left(\bigcup_{B \in A_\epsilon^n} B \right) \\ &= \sum_{B \in A_\epsilon^n} \mu(B) \\ &> |A_\epsilon^n| \, D^{-n(h+\epsilon)}. \end{aligned}$$

Esto implica que,

$$|A_\epsilon^n| < D^{n(h+\epsilon)}.$$

Asimismo, por la Propiedad de la Equipartición Asintótica y por la observación anterior, se tiene lo siguiente,

$$\begin{aligned} 1 - \epsilon < \mu & \left(\bigcup_{B \in A_\epsilon^n} B \right) \\ &= \sum_{B \in A_\epsilon^n} \mu(B) \\ &< |A_\epsilon^n| \, D^{-n(h-\epsilon)}. \end{aligned}$$

Esto implica que,

$$(1-\epsilon)D^{n(h-\epsilon)} < |A_\epsilon^n|.$$

□

5.2. Codificación de Datos

Consideremos una *fuente de información* o un *generador de mensajes*, es decir, un generador de sucesiones de símbolos. Consideremos un alfabeto $\mathcal{A}_k = \{s_1, s_2, \cdots, s_k\}$ de k símbolos y consideremos el conjunto de sucesiones infinitas de símbolos de del alfabeto $\mathcal{A}_k^{\mathbb{N}} = \Sigma_k^+$.

Así también, se van a considerar los mensajes de longitud finita. Supongamos que $\mathcal{A}_k^n$ denota el conjunto de palabras de longitud n de símbolos en $\mathcal{A}_k$. Por ejemplo, un periódico podría ser una fuente de información discreta. El conjunto de todas las palabras de longitud finita se denota como sigue,

$$\mathcal{A}^* = \bigcup_{n=1}^{\infty} \mathcal{A}_k^n.$$

Los mensajes contenidos en la fuente de información son recibidos por un codificador que los transforma en sucesiones de palabras codificadas. Las palabras codificadas son palabras en otro alfabeto que tienen correspondencia con las palabras de la fuente de información por medio del codificador. Así pues, un ***código*** es una función $C : \mathcal{A}_k \to \mathcal{B}_m^*$, donde $\mathcal{A}_k$ y $\mathcal{B}_m$ son alfabetos con k y m elementos respectivamente. Dado un alfabeto $\mathcal{A}_k$ decimos que un ***código binario*** es cualquier función $C : \mathcal{A} \to \{0,1\}^*$ es decir, una función que a cada símbolo en el alfabeto le asigna una palabra que consiste de 0's y 1's. Asimismo, para cada $s \in \mathcal{A}_k$, decimos que $C(s)$ una ***palabra codificada***. La longitud de una palabra codificada $l_C(s)$ consiste en la cantidad de símbolos en $\{0,1\}$ que tiene la palabra $C(s)$.

Ahora bien, decimos que un código binario $C : \mathcal{A}_k \to \{0,1\}^*$ es un ***código no singular*** cuando a diferentes símbolos en $\mathcal{A}_k$ les corresponden diferentes palabras código, es decir, un código C es singular cuando es inyectivo. Por ejemplo, el código Morse asigna a cada letra del abecedario una palabra compuesta con los símbolos $\cdot$ y $-$ es un código no singular.

Hasta ahora, sólo se ha mostrado la manera de codificar símbolos. Sin embargo, nos interesa codificar mensajes. Así pues, definimos una *función código* como sigue.

Definición 30. *Sea $\mathcal{A}_k$ un alfabeto con k símbolos y $C : \mathcal{A}_k \to \{0,1\}$ un código binario. Definimos una **función código** $C^* : \mathcal{A}_k^* \to \{0,1\}^*$ como sigue,*

$$C(s_{i_1} \cdots s_{i_n}) = C(s_{i_1}) \cdots C(s_{i_n}).$$

Donde $s_{i_1} \cdots s_{i_n} \in \mathcal{A}_k$

Nos interesa que cada mensaje sea decodificado de manera única. Los códigos para los cuales a cada sucesión de palabras código les corresponde una única sucesión de palabras en el código fuente reciben el nombre de ***códigos sin ambigüedad***. El siguiente ejemplo muestra un código ambiguo.

Ejemplo 17. Sea $\mathcal{A}_3 = \{a_1, a_2, a_3\}$ un alfabeto. Definimos un código binario C como sigue,

$$\begin{aligned}C(a_1) &= 1\\ C(a_2) &= 00\\ C(a_3) &= 11.\end{aligned}$$

El código que se acaba de definir no es un código sin ambigüedad, ya que no hay manera de distinguir el mensaje *aaaa* del mensaje *cc*, ya que ambos se codifican como 1111.

A continuación se muestra un ejemplo de un código sin ambigüedad.

Ejemplo 18. Sea $\mathcal{A}_3 = \{a_1, a_2, a_3\}$ un alfabeto. Definimos un código binario C como sigue,

$$\begin{aligned}C(a_1) &= 1\\ C(a_2) &= 00\\ C(a_3) &= 10.\end{aligned}$$

El código que se acaba de definir es un código sin ambigüedad. Cada palabra se puede distinguir de manera única.

5.2.1. Códigos Libres de Prefijos

Entre los códigos sin ambigüedad, se encuentran los *códigos libres de prefijos*. Sin embargo, antes de definir lo que se entenderá como un código libre de prefijos, se introducirá la noción de *prefijo*.

Sea $x \in \mathcal{A}_k^n$ una palabra de longitud n compuesta de símbolos en el alfabeto $\mathcal{A}_k$. Decimos que x es un ***prefijo*** de otra palabra x' compuesta de símbolos en $\mathcal{A}_k$ y de longitud mayor a n si los primeros n símbolos que componen a x' forman la palabra x. Por ejemplo, la palabra *abba* es un prefijo de la palabra *abbabc*.

Un código binario $C : \mathcal{A}_k \to \{0, 1\}$ es un ***código libre de prefijos*** si ninguna palabra codificada es el prefijo de otra palabra codificada.

Ejemplo 19. Sea $\mathcal{A}_3 = \{a_1, a_2, a_3\}$ un alfabeto. Definimos un código binario C como sigue,

$$\begin{aligned}C(a_1) &= 1\\ C(a_2) &= 00\\ C(a_3) &= 10.\end{aligned}$$

El código que se acaba de definir no es un código libre de prefijos, ya que 1 es un prefijo de 10. Sin embargo, si definimos un código binario C' como sigue,

$$\begin{aligned} C'(a_1) &= 0 \\ C'(a_2) &= 10 \\ C'(a_3) &= 11. \end{aligned}$$

El código resulta ser libre de prefijos, pues ninguna palabra codificada es prefijo de otra palabra codificada.

Los códigos libres de prefijos tienen dos propiedades importantes. La primera propiedad es que todo código libre de prefijos es un código sin ambigüedad ya que la condición de que ninguna palabra código sea el prefijo de otra palabra código nos asegura que si dos palabras código son diferentes, entonces corresponden a dos palabras fuente diferente. La segunda propiedad es que los códigos libres de prefijos son fácilmente decodificables. En efecto, dado que los códigos sin ambigüedad son inyectivos, entonces son invertibles y dado que las palabras codificadas no son prefijos de alguna otra palabra codificada, se puede encontrar de manera casi inmediata la palabra de la fuente de la cual provienen.

5.2.2. Arboles Asociados a un Código

Para estudiar a los códigos libres de prefijos, les asociaremos un *árbol*. Los árboles son *gráficas* que empiezan desde un *nodo raíz*. Cada nodo en una gráfica es un *nodo hoja*, o bien, es un *nodo interior*. Un nodo interior que tiene una o mas ramificaciones es llamado *padre* de las ramificaciones. Un nodo hoja es un nodo sin ramificaciones. La *aridad de un nodo* no es otra cosa que el número de ramificaciones.

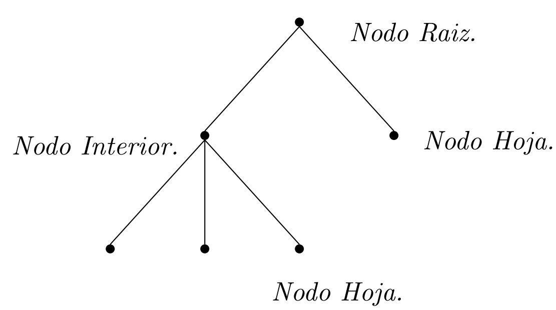
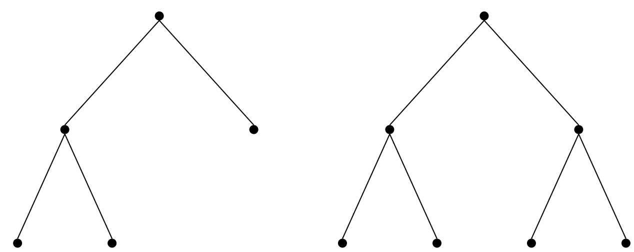


Figura 5.1: Partes de un árbol.

Se suele dibujar a los árboles poniendo la raíz hasta arriba y las hojas hasta abajo. La *profundidad de un nodo* es el número de arcos que van desde la raíz hasta el nodo. La *profundidad de un árbol* es el máximo del número de arcos que van desde la raíz hasta un nodo hoja. Finalmente, se dice que un nodo n

cubre a otro nodo m si la trayectoria que va de la raíz a m cubre la trayectoria que va de la raíz a n.

Así también, un *árbol de aridad* n es un árbol en el que todo nodo interior tiene exactamente n ramificaciones. Un *árbol completo* es un árbol de aridad n en el cual todos los nodos hojas tienen la misma profundidad.



(a) Arbol no completo. (b) Árbol Completo.

Figura 5.2: Árbol de aridad 2.

Un árbol lleno de aridad n y profundidad $d \geq 0$ tiene la propiedad de que a cada nodo a una profundidad δ $(0 \leq \delta \leq d)$ le caben exactamente,

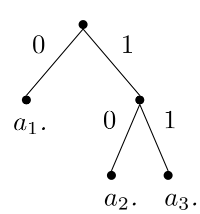
$$n^{d-\delta}, \tag{5.1}$$

hojas.

La característica principal de los códigos libres de prefijos es que una vez que se le haya asignado una sucesión finita de símbolos de un alfabeto $\mathcal{B}_m$ a cada símbolo en un alfabeto $\mathcal{A}_k$, entonces ningún otro código debe empezar con ese "patrón" o esa misma sucesión de símbolos. Ningún patrón puede ser prefijo de algún otro código.

Dados los alfabetos $\mathcal{A}_k$, $\mathcal{B}_m$ y un código libre de prefijos $C : \mathcal{A}_k \to \mathcal{B}_m^*$ le asociamos un árbol de aridad m tal que a cada palabra codificada le corresponda una sucesión de etiquetas de una trayectoria que va de la raíz a la hoja.

Ejemplo 20. En el ejemplo 19 se mostró un alfabeto $\mathcal{A}_3 = \{a_1, a_2, a_3\}$ y un código libre de prefijos C' definido de la siguiente manera: $C'(a_1) = 0$, $C'(a_2) = 10$ y $C'(a_3) = 11$. El árbol asociado al código C' es el siguiente:



5.2.3. Desigualdad de Kraft

El siguiente resultado se conoce como la *desigualdad de Kraft*. La desigualdad de Kraft nos da condiciones necesarias y suficientes para la existencia de un código libre de prefijos.

Teorema 34 (Desigualdad de Kraft)**.** *Sean* $\mathcal{A}_k = \{a_1, a_2, \cdots, a_k\}$*,* $\mathcal{B}_k = \{b_1, b_2, \cdots, b_m\}$ *dos alfabetos de* k *y* m *símbolos respectivamente. Sea* $C : \mathcal{A}_k \to \mathcal{B}_m^*$ *un código libre de prefijos. Si* $l_1, l_2, \cdots, l_k$ *son las longitudes de las palabras codificadas, entonces se satisface la siguiente desigualdad,*

$$\sum_{i=1}^{k} m^{-l_i} \leq 1.$$

Conversamente, dado un conjunto de longitudes que satisfacen la desigualdad anterior, se tiene que existe un código libre de prefijos cuyas palabras codificadas tienen estas longitudes.

La igualdad de da cuando el código libre de prefijos es completo.

Demostración. Por hipótesis $C : \mathcal{A}_k \to \mathcal{B}_m^*$ es un código libre de prefijos, entonces se le puede asociar un árbol de aridad k cuyas palabras codificadas tienen longitudes $l_1, l_2, \cdots, l_k$. Consideramos,

$$L := \max_{1 \leq i \leq k} l_i + 1.$$

Como el código C es un código libre de prefijos, entonces se tiene que ningún nodo que corresponde a una palabra codificada puede estar debajo de otro nodo que corresponde a otra palabra codificada. Así pues, por la observación (5.1) se tiene que una palabra codificada en el nivel l_i tiene a lo mas,

$$m^{L-l_i},$$

descendientes en el nivel L. Ahora bien, consideremos el código completo. Entonces se tiene lo siguiente,

$$\begin{aligned} \sum_{i=1}^{k} m^{L-l_i} &= \sum_{i=1}^{k} m^L m^{-l_i} \\ &= m^L \sum_{i=1}^{k} m^{-l_i} \end{aligned}$$

Como el árbol de aridad n asociado al código libre de prefijos puede tener a lo mas m^L hojas, entonces,

$$m^L \sum_{i=1}^{k} m^{-l_i} \leq m^L.$$

Así pues,

$$\sum_{i=1}^{k} m^{-l_i} \leq 1.$$

Ahora bien, cuando el código es completo, entonces todas las hojas corresponden a una palabra codificada. Por lo tanto,

$$m^L \sum_{i=1}^{k} m^{-l_i} = m^L,$$

es decir,

$$\sum_{i=1}^{k} m^{-l_i} = 1.$$

Para probar el recíproco, supongamos que $l_1, l_2, \cdots, l_k$ son enteros positivos tales que,

$$\sum_{i=1}^{k} m^{-l_i} \leq 1.$$

Consideremos a $L = \max_{1 \leq i \leq k} l_i$. y supongamos que n_j es el número de los l_i tales que son iguales a j, donde $1 \leq j \leq L$. Así pues,

$$\sum_{j=1}^{L} n_j m^{-j} = \sum_{i=1}^{k} m^{-l_i} \leq 1.$$

Entonces,

$$\sum_{j=1}^{L} n_j m^{-j} = \sum_{j=1}^{L-1} n_j m^{-j} + n_L m^{-L}.$$

Esto implica que,

$$n_L m^{-L} \leq 1 - \sum_{j=1}^{L-1} n_j m^{-j}$$

$$n_L \leq m^L - m^L \sum_{j=1}^{L-1} m^{-j}$$

$$= m^L - \sum_{j=1}^{L-1} m^{L-j}.$$

Análogamente,

$$n_{L-1} \leq m^{L-1} - \sum_{j=1}^{L-2} n_j m^{L-j-1}.$$

En general, para cada k tal que $0 \leq k \leq L-1$, se satisface,

$$n_{L-k} \leq m^{L-k} - \sum_{j=1}^{L-k-1} n_j m^{L-j-k}.$$

Las desigualdades anteriores nos garantizan que haya suficientes nodos para asignar a elemento del alfabeto $\mathcal{A}_k$ una palabra codificada procediendo de la siguiente manera. Ordenamos las longitudes $l_1, l_2, \cdots l_k$ de menor a mayor y le vamos asignando a cada una una rama del árbol hasta agotar todas las k longitudes. $\square$

A continuación veremos un ejemplo para ilustrar como funciona el algoritmo descrito en la prueba de la Desigualdad de Kraft.

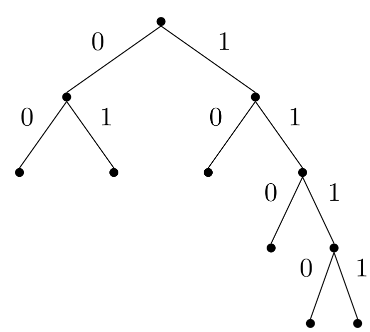
Ejemplo 21. Consideremos las siguientes longitudes:

$$\begin{aligned} l_1 &= 2 \\ l_2 &= 2 \\ l_3 &= 2 \\ l_4 &= 3 \\ l_5 &= 4. \end{aligned}$$

Para saber si existe un código binario cuyas palabras codificadas tienen las longitudes anteriores, se debe verificar que satisfaga la Desigualdad de Kraft.

$$\sum_{i=1}^{5} 2^{-l_i} = 3 \cdot \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^4} = \frac{15}{16} \leq 1.$$

Siguiendo el algoritmo descrito en la prueba, tenemos que, las palabras se codi-



fican como 00, 01, 10, 110, 1110.

Como observación final acerca de la desigualdad de Kraft, cabe mencionar que el teorema únicamente nos dice cuando podría existir un código libre de prefijos. Sin embargo, no nos dice si un código dado es un código libre de prefijo. Por ejemplo, el código $\{1, 00, 10\}$ no es un código libre de prefijo. Sin embargo, se tiene $2^{-1} + 2^{-2} + 2^{-2} = 1$. La desigualdad de Kraft sólo afirma que existe un código libre de prefijos para esas longitudes.

5.3. Compresión Optima

En la sección anterior se mostró que cualquier código que satisfaga la desigualdad de Kraft es un código libre de prefijos y nos sirve para codificar mensajes de un alfabeto dado $\mathcal{A}_k$. Sin embargo, el problema al que nos enfrentamos en esta sección es el de encontrar códigos libres de prefijos *óptimos*, en el sentido de que la codificación obtenida de una "palabra" dada tenga la longitud mínima.

La idea básica para obtener una codificación óptima de una cadena de símbolos de un alfabeto dado, consiste en asignar palabras cortas a los símbolos que aparezcan con mayor frecuencia.

Ejemplo 22. Consideremos un alfabeto $\mathcal{A}_5 = \{a_1, a_2, a_3, a_4, a_5\}$ y supongamos que queremos codificar la palabra,

$$a_5a_3a_5a_5a_3a_5a_2a_4a_2a_1a_3a_5.$$

Proponemos dos codificaciones libres de prefijos,

a_1	1	00001
a_2	01	001
a_3	001	01
a_4	0001	0001
a_5	00001	1

En el primer caso, la palabra se codifica como,

$$0000100100001000010010000101000101100100001.$$

Con el segundo código, la palabra se codifica como sigue,

$$10111011001000100100001011.$$

Mientras que con el primer código la palabra codificada consta de 43 símbolos, con el segundo código consta de 26 símbolos. Por lo tanto, la segunda codificación es mas eficiente que la primera. Aunque se utilizaron las mismas palabras para codificar, la diferencia consistió en haber asignado las palabras más cortas a los símbolos con mayor frecuencia.

Al asignar palabras cortas, dado un código libre de prefijos, a los símbolos con mayor frecuencia en una cadena se obtienen palabras codificadas cortas. En la siguiente sección se mostrará un algoritmo para encontrar las codificaciones óptimas. En el resto de esta sección vamos a calcular la tasa máxima de compresión que puede admitir una cadena de caracteres. Como ya se puede anticipar, la tasa de compresión esta en función de la "redundancia" presente en la cadena de símbolos y por ende está en función de la entropía o cantidad de información.

Dado un alfabeto $\mathcal{A}_k = \{a_1, a_2, \cdots, a_k\}$ y una cadena de símbolos de longitud r, definimos la probabilidad p_i asociada a un símbolo a_i como sigue,

$$p_i = Pr(a_i) \qquad \forall i \in \{1, 2, \cdots, k\}.$$

Donde $Pr(a_i)$ es la función de probabilidad que se calcula contando el número de veces que aparece el símbolo a_i en la cadena y dividiéndolo entre la longitud de la cadena. A continuación se define lo que se entenderá como la longitud promedio de un código.

Definición 31. *Sean $\mathcal{A}_k = \{a_1, a_2, \cdots, a_k\}$, $\mathcal{B}_m = \{b_1, b_2, \cdots, b_m\}$ dos alfabetos y $C : \mathcal{A} \to \mathcal{B}_m$ un código. $l_C : \mathcal{A}_k \to \mathbb{R}$ denota la* ***longitud de cada palabra codificada****. Definimos la* ***longitud promedio*** *del código C como,*

$$L(C) := \sum_{i=1}^{k} p_i l_C(a_i).$$

Donde $p_i)$ es la probabilidad asociada a cada símbolo $a_i \in \mathcal{A}_k$.

El problema de optimizar la tasa de compresión de una cadena de símbolos dada se puede ver como el problema de minimizar la longitud promedio de un código. Hay varias maneras de minimizar la longitud promedio. Por ejemplo, [8] encuentra que la longitud esperada coincide con la entropía de la fuente h utilizando el método de multiplicadores de Lagrange. Sin embargo, en este trabajo se verá que la Propiedad de la Equipartición Asintótica muestra de manera directa que un código óptimo en promedio codifica cada palabra con una longitud h.

Sea $\mathcal{A}_k = \{a_1, a_2, \cdots, a_k\}$ un alfabeto. Consideramos el espacio de sucesiones laterales Σ_k^+, la transformación corrimiento $\sigma : \Sigma_k^+ \to \Sigma_k^+$ y la medida de Bernoulli. Asimismo, consideramos la partición medible que consiste en los cilindros de longitud 1 $\xi = \{C_{a_1}, C_{a_2}, \cdots, C_{a_k}\}$. En la sección 4.4.1 se mostró que la partición ξ es una partición generadora, de manera que la entropía de la fuente $h = h(\sigma)$ coincide con la entropía con respecto a la partición ξ. Más aún,

$$\xi_n = \bigvee_{j=0}^{n} \sigma^{-j}\xi = \{C_{i_0 \cdots i_n} : i_0, \cdots, i_n \in \mathcal{A}_k\}.$$

Cada palabra de longitud n se puede ver como un cilindro $C_{i_0 \cdots i_n}$ que es parte de la partición ξ_n. Por la Propiedad de la Equipartición Asintótica, sabemos que podemos ver a los elementos de la partición ξ_n como la unión de conjuntos típicos A_ϵ^n y conjuntos no típicos. En la sección 5.1 se probó que una de las características principales de los conjuntos típicos es que son conjuntos con alta probabilidad, es decir, su medida es casi 1,

$$1 - \epsilon \leq \mu(A_\epsilon^n) \leq 1.$$

Mientras que la medida de su complemento es menor a ϵ y por lo tanto, es despreciable. Asimismo se probó que

$$(1 - \epsilon)D^{n(h-\epsilon)} < |A_\epsilon^n| < D^{n(h+\epsilon)}.$$

Es decir, hay aproximadamente $n(h+\epsilon)+1$ símbolos en un palabra codificada. Para un código C, la longitud promedio es la siguiente:

$$\begin{aligned} L(C) &= \sum_{i_0,\cdots,i_n} \mu(C_{i_0,\cdots,i_n}) l_C(xi_0,\cdots,x_{i_n}) \\ &= \sum_{i_0,\cdots,i_n} \mu(C_{i_0,\cdots,i_n})(n(h+\epsilon)+1) \\ &= \mu(A^n_\epsilon)(n(h+\epsilon)+1) \end{aligned}$$

Pero,

$$(1-\epsilon)(n(h+\epsilon)+1) \leq (n(h+\epsilon)+1)\mu(A^n_\epsilon) \leq (n(h+\epsilon)+1).$$

Es decir,

$$(1-\epsilon)(n(h+\epsilon)+1) \leq L(C) \leq (n(h+\epsilon)+1).$$

Dividiendo entre n y haciendo tender n a infinito se obtiene,

$$h+(1-h)\epsilon \leq \lim_{n\to\infty} \frac{L(C)}{n} \leq h+\epsilon.$$

Como ϵ es arbitraria, se concluye,

$$\lim_{n\to\infty} \frac{L(C)}{n} = h.$$

Por lo tanto, podemos representar palabras en $\mathcal{A}_k$ usando en promedio cadenas de longitud h por cada símbolo en $\mathcal{A}_k$. Una palabra de longitud n tiene en promedio una longitud nh.

5.4. Algoritmo de Huffman

En esta sección se presenta el *algoritmo de Huffman*, que es uno de los más utilizados actualmente para construir códigos óptimos. Primero se mostrará como funciona el algoritmo, veremos un ejemplo y posteriormente, probaremos que efectivamente es un algoritmo óptimo.

Consideremos un alfabeto $\mathcal{A}_k$ y la distribución de probabilidades asociada a cada símbolo en $\mathcal{A}_k$. El algoritmo de Huffman comienza construyendo una lista de todos los elementos de $\mathcal{A}_k$ en el orden descendiente de sus probabilidades. Entonces se construye un árbol con un símbolo en cada hoja desde abajo hasta arriba. En cada paso se seleccionan los dos símbolos con probabilidades más pequeñas, se suman, se reemplazan con un símbolo auxiliar que representa los dos símbolos originales y se añade a un escalón del árbol parcial. El proceso termina cuando la lista se reduce a un sólo símbolo auxiliar. Veamos un ejemplo.

Ejemplo 23. Consideremos el alfabeto $\mathcal{A}_k = \{a_1, a_2, \cdots, a_8\}$ con sus respectivas probabilidades,

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
0,05	0,05	0,1	0,1	0,15	0,15	0,2	0,2.

Observamos que las probabilidades ya están en orden descendiente. De manera que se procede a sumar las probabilidades de a_1 y de a_2 y reemplazar por el símbolo a_{12}. Entonces queda,

$$\begin{array}{lllllll} a_{12} & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 \\ 0{,}1 & 0{,}1 & 0{,}1 & 0{,}15 & 0{,}15 & 0{,}2 & 0{,}2. \end{array}$$

Como de nuevo ya están en orden descendiente, se suman las probabilidades de a_{12} y las de a_3 y se reemplazan dichos símbolos por a_{123}.

$$\begin{array}{llllll} a_{123} & a_4 & a_5 & a_6 & a_7 & a_8 \\ 0{,}2 & 0{,}1 & 0{,}15 & 0{,}15 & 0{,}2 & 0{,}2. \end{array}$$

Después de reordenar, los símbolos con menores probabilidades asociadas son a_4 y a_5. Después de sumar y reemplazar por a_{45} se obtiene,

$$\begin{array}{lllll} a_{45} & a_6 & a_{123} & a_7 & a_8 \\ 0{,}25 & 0{,}15 & 0{,}2 & 0{,}2 & 0{,}2. \end{array}$$

De nueva cuenta se vuelve a ordenar, se suman las probabilidades de los símbolos a_6 y a_{123}. Reemplazamos por el símbolo a_{1236}.

$$\begin{array}{llll} a_{1236} & a_7 & a_8 & a_{45} \\ 0{,}35 & 0{,}2 & 0{,}2 & 0{,}25. \end{array}$$

En esta ocasión las probabilidades que se deben sumar son las correspondientes a a_7 y a_8.

$$\begin{array}{lll} a_{78} & a_{45} & a_{1236} \\ 0{,}4 & 0{,}25 & 0{,}35. \end{array}$$

Finalmente, repitiendo una vez más el procedimiento se obtiene,

$$\begin{array}{ll} a_{451236} & a_{78} \\ 0{,}6 & 0{,}4. \end{array}$$

El árbol construido mediante este procedimiento se muestra en la figura 5.3 Para asignar los códigos, simplemente asignamos 0's y 1's a las ramas izquierda y derecha respectivamente de arriba para abajo como se muestra en la figura 5.4. Por lo tanto, siguiendo el algoritmo de Huffman, los símbolos del alfabeto $\mathcal{A}_k$ se codifican como sigue:

$$\begin{array}{ll} a_1 & 00000 \\ a_2 & 00001 \\ a_3 & 0001 \\ a_4 & 010 \\ a_5 & 011 \\ a_6 & 001 \\ a_7 & 10 \\ a_8 & 11. \end{array}$$

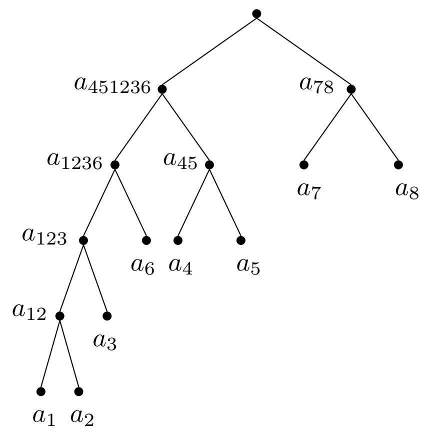


Figura 5.3: Árbol contruído con el algoritmo de Huffman.

Cabe observar que los códigos asignados por el algoritmo de Huffman no son únicos, ya que en la mayoría de las ocasiones hay más de un símbolo que comparte la menor probabilidad asociada. Por ejemplo, en el ejemplo anterior el símbolo a_{12} se pudo haber combinado tanto con a_3 como con a_4. La elección de combinarlo con a_3 es completamente arbitraria.

Los algoritmos de Huffman se suelen utilizar principalmente para construir códigos binarios. Sin embargo, cabe mencionar que el mismo procedimiento permite construir cualesquiera otros códigos. Por ejemplo, para construir códigos ternarios. En lugar de combinar los dos símbolos con probabilidades más pequeñas, combinamos los tres símbolos con probabilidades más pequeñas y continuamos el procedimiento de la misma manera.

5.4.1. Códigos Óptimos y Códigos de Huffman

En esta sección probaremos que el algoritmo descrito en la sección anterior nos permite construir códigos óptimos, es decir, códigos en los cuales la longitud esperada es mínima. Los códigos construidos a partir del algoritmo de Huffman reciben el nombre de ***códigos de Huffman***. Para probar que los códigos de Huffman son óptimos, probaremos primero un lema que nos asegura que cada paso en el algoritmo de Huffman permite ir construyendo códigos óptimos. Posteriormente se probará por inducción el resultado principal.

Para desarrollar el algoritmo de Huffman, considerábamos una distribución de probabilidades asociadas a los símbolos de un alfabeto y nos fijábamos en las dos menores. Al sumar las dos probabilidades menores se construía una nueva distribución con un elemento menos que la distribución original. La nueva distribución recibe el nombre de ***reducción de Huffman***.

Lema 22. *Sea $\mathcal{A}_k = \{a_1, a_2, \cdots, a_k\}$ un alfabeto. Dado un código óptimo para*

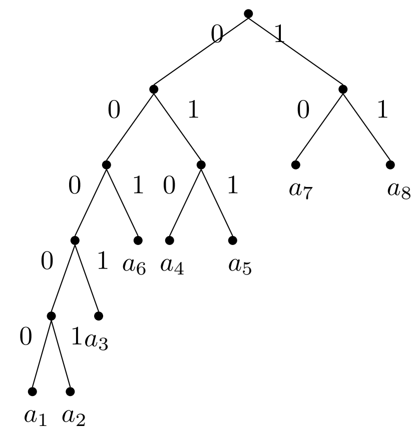


Figura 5.4: Árbol con códigos.

una distribución de probabilidades P, se puede construir un código óptimo para la reducción de Huffman siguiendo el primer paso del algoritmo de Huffman.

Demostración. Sea $P = \{p_1, p_2, \cdots, p_k\}$ una distribución de probabilidades tal que $p_1 \leq p_2 \leq \cdots \leq p_k$. Consideramos la reducción de Huffman Q, es decir,

$$Q = \{p_1 + p_2, p_3, \cdots, p_k\}.$$

Supongamos que C_P^* y C_Q^* son códigos óptimos para las distribuciones P y Q respectivamente. A partir del código óptimo C_Q^* podemos construir una extensión para k elementos. Primero tomamos $C_Q^*(a_{12})$ que es la palabra codificada que corresponde al primer elemento $p_{12} = p_1 + p_2$ de la distribución Q y definimos un codigo C_P como sigue:

$$\begin{aligned} C_P(a_1) &= C_Q^*(a_{12})0 \\ C_P(a_2) &= C_Q^*(a_{12})1 \\ C_P(a_i) &= C_Q^*(a_i) \qquad \forall i \in \{3, 4, \cdots, k\}. \end{aligned}$$

Por la manera en que se definieron los códigos C_P y C_Q^* podemos hacer dos observaciones importantes,

- $l_{C_P}(a_1) = l_{C_P}(a_2) = l_{C_Q^*}(a_{12}) + 1.$

- $L((C_Q^*) = p_{12} l_{C_Q^*}(a_{12}) + \sum_{i=3}^{k} p_i l_{C_Q^*}(a_i) = p_{12} l_{C_Q^*}(a_{12}) + \sum_{i=3}^{k} p_i l_{C_P}(a_i).$

Así pues, con base en las observaciones anteriores calculamos la longitud esperada $L(C_P)$.

$$
\begin{aligned}
L(C_P) &= \sum_{i=1}^{k} p_i l_{C_P}(a_i) \\
&= \sum_{i=3}^{k} p_i l_{C_P}(a_i) + p_1 l_{C_P}(a_1) + p_2 l_{C_P}(a_2) \\
&= \sum_{i=3}^{k} p_i l_{C_P}(a_i) + p_1 (l_{C_Q^*}(a_{12}) + 1) + p_2 (l_{C_Q^*}(a_{12}) + 1) \\
&= \sum_{i=3}^{k} p_i l_{C_P}(a_i) + (p_1 + p_2) l_{C_Q^*}(a_{12}) + p_1 + p_2 \\
&= \sum_{i=3}^{k} p_i l_{C_P}(a_i) + (p_{12}) l_{C_Q^*}(a_{12}) + p_1 + p_2 \\
&= L((C_Q^*) + p_1 + p_2.
\end{aligned}
$$

Supongamos que C_P^* es un código óptimo para la distribución P y que C_Q es un código libre de prefijos para la distribución Q. Como $p_1 \leq p_2 \leq \cdots \leq p_k$, entonces $l_{C_P^*}(a_k) \leq l_{C_P^*}(a_{k-1}) \leq \cdots \leq l_{C_P^*}(a_1)$. Más aun, podemos afirmar que $l_{C_P^*}(a_1) = l_{C_P^*}(a_2)$. En efecto, si no tuvieran la misma longitud, podríamos remover el último símbolo de la palabra mas larga y aun evitar que la palabra tenga prefijos. Esto implica que se podría alcanzar una longitud esperada menor, lo cual es imposible.

Así también, podemos suponer sin pérdida de generalidad que las dos palabras codificadas más largas $C_P^*(a_1)$ y $C_P^*(a_2)$ difieren únicamente en el último símbolo. Los códigos óptimos no satisfacen necesariamente esta propiedad. Sin embargo, dado un código óptimo, siempre se puede encontrar otro código óptimo que sí satisfaga la propiedad simplemente reacomodando. Basta intercambiar las palabras codificadas mas grande de manera que los dos símbolos de probabilidad menor estén asociados de tal manera que sus palabras codificadas sólo difieran en el último símbolo.

Con base en las observaciones anteriores construimos un código C_Q para la distribución Q. Dado que $C_P^*(a_1)$ y $C_P^*(a_2)$ difieren sólo en el último símbolo, entonces definimos $C_Q(a_{12})$ como $C_P^*(a_1)$ menos el último símbolo y para cada $i \in \{3, 4, \cdots, k\}$ definimos $C_Q(a_i) = C_P^*(a_i)$.

Por la manera en que se definió el código C_Q, podemos hacer dos observaciones importantes,

- $l_{C_Q}(a_{12}) = l_{C_P^*}(a_1) - 1$
- $l_{C_Q}(a_i) = l_{C_P^*}(a_i) \qquad \forall i \in \{3, 4, \cdots, k\}$.

Calculamos la longitud esperada para el código C_Q,

$$\begin{aligned}
L(C_Q) &= \sum_{i=3}^{k} p_i l_{C_Q}(a_i) + p_{12} l_{C_Q}(a_{12}) \\
&= \sum_{i=3}^{k} p_i l_{C_Q}(a_i) + p_{12}(l_{C_P^*}(a_1) - 1) \\
&= \sum_{i=3}^{k} p_i l_{C_Q}(a_i) + p_1 l_{C_P^*}(a_1) - p_1 + p_2 l_{C_P^*}(a_2) - p_2) \\
&= \sum_{i=3}^{k} p_i l_{C_P^*}(a_i) + p_1 l_{C_P^*}(a_1) - p_1 + p_2 l_{C_P^*}(a_2) - p_2) \\
&= \sum_{i=1}^{k} p_i l_{C_P^*}(a_i) - p_1 - p_2 \\
&= L(C_P^*) - p_1 - p_2.
\end{aligned}$$

Sumando las ecuaciones obtenidas,

$$L(C_Q) + L(C_P) = L(C_Q^*) + L(C_P^*),$$

o bien,

$$L(C_Q) - L(C_Q^*) + L(C_P) - L(C_P^*) = 0.$$

Como C_Q^*, C_P^* son códigos óptimos, se sigue que $L(C_Q^*) \leq L(C_Q)$ y que $L(C_P^*) \leq L(C_P)$. Esto implica,

$$L(C_Q) - L(C_Q^*) \geq 0,$$

$$L(C_P) - L(C_P^*) \geq 0.$$

Como la suma de dos números positivos da 0, se sigue que cada uno de los sumando debe ser 0. Así pues, se obtiene que,

$$L(C_Q) = L(C_Q^*),$$

$$L(C_P) = L(C_P^*).$$

Por lo tanto, dado un código óptimo, el primer paso del algoritmo de Huffman nos da a su vez un código óptimo para la reducción de Huffman. □

A continuación se probará que los algoritmos de Huffman nos permiten construir códigos óptimos.

Teorema 35. *Sea $\mathcal{A}_k = \{a_1, a_2, \cdots, a_k\}$ un alfabeto. Si $C_H : \mathcal{A}_k \to \{0,1\}^*$ es un código de Huffman y $C : \mathcal{A}_k \to \{0,1\}^*$ es un código libre de prefijos, entonces $L(C_H) \leq L(C)$.*

Demostración. Por inducción sobre la cardinalidad del alfabeto. Para probar la base de la inducción, supongamos que $k = 2$. Al asignar el 1 a un símbolo y 2 al otro se obtiene un código óptimo.

Por hipótesis de inducción, supongamos que para un alfabeto de k símbolos se tiene un código óptimo y veamos que podemos construir un código óptimo para un alfabeto de $k+1$ símbolos. Siguiendo el algoritmo de Huffman, sumamos las dos probabilidades menores y reemplazamos los símbolos correspondientes por otro símbolo. Observamos que la distribución que queda no es otra cosa que la reducción de Huffman. Por la hipótesis de inducción, se tiene que existe un código óptimo para el nuevo alfabeto y por el Lema anterior, a partir del código óptimo para el alfabeto de k símbolos, podemos construir un código óptimo para el alfabeto de $k + 1$ símbolos. □

El teorema anterior se probó para códigos de Huffman binarios. Sin embargo, la prueba tanto del lema como del teorema se puede extender de manera análoga para establecer que los códigos de Huffman de aridad D son códigos óptimos.

5.4.2. Programas en lenguaje C

En esta sección se muestra un programa realizado en lenguaje C que codifica un texto utilizando el algoritmo de Huffmann. El programa utiliza principalmente estructuras dinámicas de datos, como son las "colas" y los "árboles binarios".

El programa utiliza tres estructuras. Una llamada "código", otra llamada "queueNode" y una llamada "textNode". La estructura textNode consta de un caracter y un apuntador, se utiliza para almacenar cada caracter que se escriba en el teclado en una cola. La cola que se crea con las estructuras textNode nos servirán para escribir el texto codificado. La estructura queueNode es la más importante del programa, es la que se ocupa para construir el árbol de Huffmann; almacena un caracter, un número entero y cinco apuntadores:

1. code.
2. nextPtr.
3. suivPtr.
4. leftPtr.
5. rightPtr

El programa recibe información del teclado. Cada caracter que se escribe se almacena en la variable de caracter y se van referenciando en una estructura de tipo cola por medio del apuntador "nextPtr" y el apuntador "suivPtr". Sin embargo, a diferencia de lo que se hace con los nodos de tipo textPtr, en el este caso, para cada nueva letra que se escriba, se llama a la función "buscar". La función buscar recorre cada nodo de la cola que se ha ido construyendo y compara el nuevo caracter con los anteriores para ver si ya se escribió anteriormente. Si se escribió anteriormente, únicamente suma un uno a la variable de tipo entero del nodo correspondiente. Si no, crea un nuevo nodo al que le corresponde el nuevo caracter. De esta manera, se puede ir construyendo una tabla de frecuencias de los caracteres.

Una vez que hemos construido la cola correspondiente a cada caracter distinto con su respectivo número de ocurrencias, construimos el árbol de Huffmann mediante el mismo procedimiento que describimos anteriormente, es decir, utilizamos una función "Order" que lo que hace es ordenar de menor a mayor los nodos según su número de ocurrencias en el texto. Cabe mencionar que el ordenamiento se hace únicamente con los apuntadores nextPtr, mientras que se deja inalterado el orden descrito por los apuntadores suivPtr. Después de ordenar la cola de menor a mayor, se crea un nuevo nodo que tiene asociado un número que consiste en la suma de las ocurrencias de los dos últimos. Sustituimos los últimos dos nodos de la cola por este nuevo nodo y hacemos que el apuntador leftPtr referencié al nodo que antiguamente ocupaba el último lugar, mientras que el apuntador rightPtr referencía al nodo que antiguamente ocupaba el penúltimo lugar. Recordamos que la estructura queueNode contenía también un apuntador de tipo código. El apuntador que tiene asociado el nodo que quedo a la izquierda del árbol le asociamos un "0", mientras que al que quedo a la derecha le asociamos un "1". La idea es que al irse creando el árbol de Huffmann podamos ir haciendo una cola por cada ruta desde abajo hasta la raíz, de manera que dicha ruta nos vaya dando el código asociado a cada caracter distinto que se escribió en el teclado.

Finalmente, el programa despliega los resultados. Primero muestra una tabla donde se indica el caracter, el número de ocurrencias, la longitud que tiene el código que está asociado y por último el código que le corresponde. Después nos muestra la longitud de la cadena de caracteres, el valor de la entropía para ese texto en particular y al final despliega el texto codificado. A continuación se muestra el código del programa.

```
# include<string.h>
# include<stdio.h>
# include<stdlib.h>
# include<math.h>

struct codigo{
    char codeData;
    struct codigo *proxPtr;
};

struct queueNode{
    char data;
    int conta;
    struct codigo *code;
    struct queueNode *nextPtr;
    struct queueNode *suivPtr;
    struct queueNode *leftPtr;
```

```
    struct queueNode *rightPtr;
};

struct textNode{
    char data;
    struct textNode *nextPtr;
};

typedef struct textNode TEXTNODE;
typedef TEXTNODE *TEXTNODEPTR;

typedef struct codigo CODENODE;
typedef CODENODE *CODENODEPTR;

typedef struct queueNode QUEUENODE;
typedef QUEUENODE *QUEUENODEPTR;

void printQueue(QUEUENODEPTR);
void printText(TEXTNODEPTR);
void printCode(CODENODEPTR);
void codifiedText(TEXTNODEPTR, QUEUENODEPTR);
void everyCode(CODENODEPTR, int);

int isEmpty(QUEUENODEPTR);
int isEmptyTx(TEXTNODEPTR);
double entropy(QUEUENODEPTR *, int);

void enqueue(QUEUENODEPTR *, QUEUENODEPTR *, char);
void enqueueText(TEXTNODEPTR *, TEXTNODEPTR *, char);

QUEUENODEPTR buscar(QUEUENODEPTR *, char);
QUEUENODEPTR Order(QUEUENODEPTR *, int);
QUEUENODEPTR HuffmanTree(QUEUENODEPTR *, int);

main() {
    QUEUENODEPTR currentPtr = NULL;
    QUEUENODEPTR headPtr = NULL, tailPtr = NULL;
    QUEUENODEPTR rootPtr = NULL;
    QUEUENODEPTR aux = NULL;

    TEXTNODEPTR headTxPtr = NULL, tailTxPtr = NULL;

    int longCad = 0;
    int NoCar = 0;
    double h = 0;
    double calc = 0;
```

```
    char item;

    printf(``************************Algoritmo de Huffmann*******
*******************\n\n'');
    printf(``Este programa da una codificación óptima por medio del
algoritmo de Huffman. '');
    printf(``A continuación escriba el texto que se desea codificar:
\n\n'');
    while ((item = getchar()) != '\n') {
       longCad++;
       enqueueText(& headTxPtr, & tailTxPtr, item);
       if(!isEmpty(headPtr)){
          if(buscar(& headPtr, item) == NULL){
             NoCar++;
             enqueue(& headPtr, & tailPtr, item);
          }
          else
             buscar(& headPtr, item)->conta ++;
       }
       else
            enqueue(& headPtr, & tailPtr, item);
    }
    printf(``\nla longitud de la cadena es:%d.\n'', longCad);
    printf(``\nEl número de caracteres distintos es:%d.\n", NoCar);

    h = entropy(&headPtr, longCad);     printf(``\nLa entropía vale:
%f \n", h);

    headPtr = Order(&headPtr, NoCar);

    aux = headPtr;
    rootPtr = HuffmanTree(& headPtr, NoCar);
    printQueue(aux);

    calc = h*longCad;
    printf(``\nLa longitud esperada de la cadena con base en la
entropía es:%f\n'',calc );

    codifiedText(headTxPtr, aux);

    printf(``\n\nFin del programa.\n\n'');

    return 0;
}

\\Cálculo de la entropía.
```

```
double entropy(QUEUENODEPTR *sPtr, int lon){
    double result = 0;
    double p;
    QUEUENODEPTR currentPtr = NULL;

    currentPtr = *sPtr;
    while(currentPtr != NULL){
       p = (double)(currentPtr->conta)/lon;
       result = p*log2(p) + result;
       currentPtr = currentPtr->nextPtr;
    }
    result = -result;
    return result;
}

\\Inserta un Nodo en la Cola.
void enqueue(QUEUENODEPTR *headPtr, QUEUENODEPTR *tailPtr, char
value){
    QUEUENODEPTR newPtr;
    CODENODEPTR codePtr;

    codePtr = (CODENODE*)malloc(sizeof(CODENODE));
    if(codePtr != NULL){
       codePtr->codeData = 'a';
       codePtr->proxPtr =NULL;
    }
    else
       printf("\nNo insertado. Memoria no disponible.\n");

    newPtr = (QUEUENODE*)malloc(sizeof(QUEUENODE));
    if(newPtr != NULL){
       newPtr->data = value;
       newPtr->conta = 1;
       newPtr->nextPtr = NULL;
       newPtr->suivPtr = NULL;
       newPtr->leftPtr = NULL;
       newPtr->rightPtr = NULL;
       newPtr->code = codePtr;

       if(isEmpty(*headPtr))
          *headPtr = newPtr;
       else{
          (*tailPtr)->nextPtr = newPtr;
          (*tailPtr)->suivPtr = newPtr;
       }
```

```
      *tailPtr = newPtr;
   }
   else
      printf(``%c No insertado. Memoria no disponible.\n'', value);
}

\\Inserta un Nodo en la Cola que almacena el texto.
void enqueueText(TEXTNODEPTR *headPtr, TEXTNODEPTR *tailPtr, char
value){
   TEXTNODEPTR newPtr;

   newPtr = (TEXTNODE*)malloc(sizeof(TEXTNODE));
   if(newPtr != NULL){
      newPtr->data = value;
      newPtr->nextPtr = NULL;

      if(isEmptyTx(*headPtr))
         *headPtr = newPtr;
      else{
         (*tailPtr)->nextPtr = newPtr;
      }
      *tailPtr = newPtr;
   }
   else
      printf(``%c No insertado. Memoria no disponible.\n'', value);
}

\\Regresa 1 si la cola está vacía y 0 si no.
int isEmptyTx(TEXTNODEPTR headPtr){
   return headPtr == NULL;
}

\\Regresa 1 si la cola está vacía y 0 si no.
int isEmpty(QUEUENODEPTR headPtr){
   return headPtr == NULL;
}

\\Imprime la Cola.
void printQueue(QUEUENODEPTR currentPtr){
   QUEUENODEPTR aux = NULL, inicializador = NULL;
   int counting = 0, i, j;

   CODENODEPTR coco = NULL;
   inicializador = currentPtr;

   if(currentPtr == NULL)
```

```
      printf(‘‘La cola está vacía.\n’’);
    else{
      printf(‘‘\nCaracter \t Ocurrencias \t Longitud \t Código.\n’’);

      while(currentPtr != NULL){
         printf(‘‘\n%c \t\t%d \t\t ’’, currentPtr->data,
currentPtr->conta);
         aux = buscar(&currentPtr, currentPtr->data);

         printCode(aux->code);
         currentPtr = currentPtr->suivPtr;
      }
    }
}

void printCode(CODENODEPTR coding){
    CODENODEPTR coco = NULL;
    int counting = 0;
    int i, j;

    coco = coding;

    while((coding)->proxPtr != NULL){
      counting++;
      coding = coding->proxPtr;
    }

    printf(‘‘%d \t \t ’’, counting);
    everyCode(coco, counting);
}

\\Imprime cada código.
void everyCode(CODENODEPTR codito, int count){
    int i,j;
    CODENODEPTR helpCode = NULL;

    if(count > 1){
      j=1;
      while(count-j >= 0){
         helpCode = codito;

      for(i=1; i <= count - j; i++)
         helpCode = helpCode->proxPtr;
         printf(‘‘%c’’, helpCode->codeData);
         j++;
      }
```

```
    }
}

\\Imprime el texto.
void printText(TEXTNODEPTR currentPtr){
    if(currentPtr == NULL)
       printf("La cola está vacía.\n");
    else{
       while(currentPtr != NULL){
          printf(‘‘%c’’, currentPtr->data);
          currentPtr = currentPtr->nextPtr;
       }
    }
}

\\imprime el texto codificado.
void codifiedText(TEXTNODEPTR textsPtr, QUEUENODEPTR queuesPtr){
    QUEUENODEPTR aux;
    CODENODEPTR coco;
    int counter;
    int longitud = 0;

    if(textsPtr == NULL)
       printf(‘‘La cola está vacía.\n’’);
    else{
       printf(‘‘\nEl texto codificado es el siguiente:\n\n’’);

       while(textsPtr != NULL){
          counter = 0;
          aux = buscar(&queuesPtr, textsPtr->data);
          coco = aux->code;

             while((aux->code)->proxPtr != NULL){
                counter++;
                aux->code = (aux->code)->proxPtr;
             }
          aux->code = coco;
          everyCode(aux->code, counter);

          longitud = longitud + counter;

          textsPtr = textsPtr->nextPtr;

       }
       printf(‘‘\n\nLa longitud de la cadena de códificada es:%d’’,
longitud);
```

```
    }
}

QUEUENODEPTR buscar(QUEUENODEPTR *sPtr, char value){
    QUEUENODEPTR previousPtr, currentPtr;

    if(value == (*sPtr)->data){
       return *sPtr;
    }
    else{
       previousPtr = *sPtr;
       currentPtr = (*sPtr)->suivPtr;

       while(currentPtr != NULL && currentPtr->data != value){
          previousPtr = currentPtr;
          currentPtr = currentPtr->suivPtr;
       }
       if(currentPtr != NULL){
          return currentPtr;
       }
    }
    return NULL;
}

\\Ordenación tipo burbuja.
QUEUENODEPTR Order(QUEUENODEPTR *sPtr, int cantida){
    QUEUENODEPTR previousPtr, currentPtr, returnPtr, tempPtr, auxPtr;
    int pass, i;

    returnPtr = *sPtr;
    previousPtr = *sPtr;
    currentPtr = (*sPtr)->nextPtr;
    for(pass = 1; pass <= cantida; pass++){ \\Número de pasadas.
       previousPtr = returnPtr;
       currentPtr = returnPtr->nextPtr;
       if(previousPtr->conta > currentPtr->conta){
          previousPtr->nextPtr = currentPtr->nextPtr;
          currentPtr->nextPtr = previousPtr;
          returnPtr = currentPtr;
       }
       else{
          for(i = 1; i <= cantida; i++){
             if(previousPtr->conta <= currentPtr->conta){

                tempPtr = previousPtr;
                previousPtr = currentPtr;
```

```
                currentPtr = currentPtr->nextPtr;

            }
            else{
                \\Intercambia los nodos.
                tempPtr->nextPtr = currentPtr;
                previousPtr->nextPtr = currentPtr->nextPtr;
                currentPtr->nextPtr = previousPtr;

                auxPtr = previousPtr;
                previousPtr = currentPtr;
                currentPtr = auxPtr;

                tempPtr = previousPtr;
                previousPtr = currentPtr;
                currentPtr = currentPtr->nextPtr;
            }
        }
    }
  }
  return returnPtr;
}

\\Aquí se construye el árbol de Huffmann.
QUEUENODEPTR HuffmanTree(QUEUENODEPTR *sPtr, int cantida){
   QUEUENODEPTR previousPtr, currentPtr, newPtr, recorrePtr, pruebaPtr;
   CODENODEPTR codePtr;
   int i;

   previousPtr = *sPtr;
   currentPtr = (*sPtr)->nextPtr;

   for(i = 1; i <= cantida; i++){
      codePtr = (CODENODE*)malloc(sizeof(CODENODE));
      if(codePtr != NULL){
         codePtr->codeData = 'a';
         codePtr->proxPtr =NULL;
      }
      else
         printf(``\nNo insertado. Memoria no disponible.\n'');

      newPtr = (QUEUENODE*)malloc(sizeof(QUEUENODE));
      newPtr->conta = previousPtr->conta + currentPtr->conta;
      newPtr->data = 'H';
      newPtr->leftPtr = previousPtr;
      newPtr->rightPtr = currentPtr;
```

```
        newPtr->nextPtr = currentPtr->nextPtr;
        newPtr->suivPtr = NULL;
        newPtr->code = codePtr;

        (previousPtr->code)->codeData = '0';
        (previousPtr->code)->proxPtr = codePtr;

        (currentPtr->code)->codeData = '1';
        (currentPtr->code)->proxPtr = codePtr;

        previousPtr = Order (& newPtr, cantida - i);
        currentPtr = previousPtr->nextPtr;

    }
    return previousPtr;
}
```

Para ilustrar el funcionamiento del programa, consideramos un extracto de un poema de Julio Herrera y Reissing que se titula "Desolación absurda". El resultado que se obtiene después de correr el programa e ingresar el texto por medio del teclado, se muestra en la figura 5.5.

¡Ven, declina tu cabeza
de honda noche delincuente
sobre mi tétrica frente,
sobre mi aciaga cabeza;
deje su indócil rareza
tu numen desolador,
que en el drama inmolador
de nuestros mundos abrazos
yo te abriré con mis brazos
un paréntesis de amor!

El programa nos da información acerca del texto. Nos da la longitud de la cadena, el número de caracteres distintos y el valor de la entropía.

Longitud de la cadena	233
Caracteres distintos	29
Entropía	4.14
Longitud esperada	966.09
Longitud de la cadena codificada	973

Ahora bien, cuando introdujimos el concepto de entropía, mencionamos que servía como una medida a la cantidad de información. Un texto tenía menor información a mayor redundancia. A continuación vamos a comparar dos textos con la misma longitud. Uno de los textos tiene mayor redundancia que el otro, de modo que la codificación del texto con mayor redundancia resultará menor que la de mayor redundancia. Para considerar textos de mayor extensión, hacemos la siguiente modificación al método *main()*. Introducimos un apuntador "fichero" de tipo FILE hacia un archivo de texto "prueba.txt" y recorremos cada caracter con la función *getc(fichero)*.

```
main() {
 QUEUENODEPTR currentPtr = NULL;
 QUEUENODEPTR headPtr = NULL, tailPtr = NULL;
 QUEUENODEPTR rootPtr = NULL;
 QUEUENODEPTR aux = NULL;

 TEXTNODEPTR headTxPtr = NULL, tailTxPtr = NULL;

 int longCad = 0;
 int NoCar = 0;
 double h = 0;
 double calc = 0;
 char item;
 FILE *fichero;

 printf("************************Algoritmo de Huffmann*******
*******************\n\n");
 printf("Este programa da una codificación óptima por medio del algoritmo
de Huffman. ");
 while (feof(fichero) == 0) {
    longCad++;
    item = getc(fichero);
    enqueueText(& headTxPtr, & tailTxPtr, item);
    if(!isEmpty(headPtr)){
       if(buscar(& headPtr, item) == NULL){
          NoCar++;
          enqueue(& headPtr, & tailPtr, item);
       }
       else
          buscar(& headPtr, item)->conta ++;
    }
    else
         enqueue(& headPtr, & tailPtr, item);
 }
 printf("\nla longitud de la cadena es: %d.\n", longCad);
 printf("\nEl número de caracteres distintos es: %d.\n", NoCar);

 h = entropy(&headPtr, longCad);      printf("\nLa entropía vale:
%f \n", h);

 headPtr = Order(&headPtr, NoCar);
```

```
    aux = headPtr;
    rootPtr = HuffmanTree(& headPtr, NoCar);
    printQueue(aux);

    calc = h*longCad;
    printf("\nLa longitud esperada de la cadena con base en la
entropía es: %f\n",calc );

    codifiedText(headTxPtr, aux);

    printf("\n\nFin del programa.\n\n");

    return 0;
}
```

El primer texto que vamos a considerar es un extracto del poema de Leopoldo Lugones titulado "Oda a la desnudez". El texto es el siguiente:

> ¡Ah! muerde con tus dientes luminosos,
> muerde en el corazón las prohibidas
> manzanas del Edén; dame tus pechos,
> cálices del ritual de nuestra misa
> de amor; dame tus uñas, dagas de oro,
> para sufrir tu posesión maldita;
> el agua de sus lágrimas culpables;
> tu beso en cuyo fondo hay una espina.
> Mira la desnudez de las estrellas;
> la noble desnudez de las bravías
> panteras de Nepal, la carne pura
> de los recién nacidos; tu divina
> desnudez que da luz como una lámpara
> de ópalo, y cuyas vírgenes primicias
> disputaré al gusano que te busca,
> para morderte con su helada encía
> el panal perfumado de tu lengua,
> tu boca, con frescuras de piscina.
> Que mis brazos rodeen tu cintura
> como dos llamas pálidas, unidas
> alrededor de una ánfora de plata
> en el incendio de una iglesia antigua.
> Que debajo mis párpados vigilen
> la sombra de tus sueños mis pupilas
> cual dos fieras leonas de basalto
> en los portales de una sala egipcia.
> Quiero que ciña una corona de oro

tu corazón, y que en tu frente lilia
caigan mis besos como muchas rosas,
y que brille tu frente de Sibila
en la gloria cirial de los altares,
como una hostia de sagrada harina;
y que triunfes, desnuda como una hostia,
en la pascua ideal de mis delicias.

El resultado que se obtiene después de ingresar el texto al programa, se expone a continuación.

Longitud de la cadena	1198
Caracteres distintos	41
Entropía	4.31
Longitud esperada	5169.73
Longitud de la cadena codificada	5215

El segundo texto, que es con el que vamos a hacer la comparación es la letra de la canción "Chilanga banda" del grupo llamado "Café Tacuva". La texto que se presenta es un texto mucho más redundante que el primer texto, de manera que aunque tengan la misma longitud, su codificación deberá más corta. El texto es el siguiente:

Ya chole chango chilango
que chafa chumba te chotas
no checa andas de tacuche
y chale con la charola

Tan choncho como una chinche
mas chueco que la fayuca
con fusca y con cachiporra
te paso andar de guarura

Mejor yo me hecho una chela
y chance enchufo una chava
chambeando de chafirete
me sobra chupe y pachanga

Si choco saco chipote
la chota no es muy molacha
chiveando a los que machucan
se va en morder su talacha

De noche caigo al congal
no manches dice la changa
al choro de teporocho

enchifla pasa la pacha

Pachuco cholos y chundos
chichinflas y malafachas
aca los chompiras rifan
y bailan tibiri tabara

Pachuco cholos y chundos
chichinflas y malafachas
aca los chompiras rifan
y bailan tibiri tabara
y bailan tibiri tabara
y bailan tibiri tabara

Mejor yo me hecho una chela
y chance enchufo una chava
chambeando de chafirete
me sobra chupe pachanga

Mi ñero mata la bacha
y canta la cucaracha
su choya vive de chochos
de chemo churro y garnachas

Pachuco cholos y chundos
chichinflas y malafachas
aca los chompiras rifan
y bailan tibiri tabara

Pachuco cholos y chundos
chichinflas y malafachas
aca los chompiras rifan
y bailan tibiri tabara

El resultado que se obtiene después de ingresar el texto al programa, se expone a continuación.

Longitud de la cadena	1193
Caracteres distintos	31
Entropía	4.07
Longitud esperada	4857.18
Longitud de la cadena codificada	4903

La entropía del primer texto fue de 4.31, mientras que del segundo fue 4.07. A pesar de que ambos textos tienen prácticamente la misma longitud, la longitud de la primera cadena codificada fue de 5215, mientras que la de la segunda fue 4903. La diferencia fue sólo de 312 caracteres.

Figura 5.5: Resultado.

Conclusiones

El problema de la compresión de datos es un problema de optimización. Dado un alfabeto, podemos definir un código de muchas maneras. Sin embargo, nos interesa encontrar aquel código tal que las palabras codificadas tengan longitud mínima.

Los códigos con los que nos interesa trabajar son los llamados códigos libres de prefijos porque, no sólo son inyectivos, sino que dada la manera en que se definen, son libres de ambigüedad. Los códigos sin ambigüedad nos permiten encontrar de manera unívoca los símbolos de los cuales provienen, de manera que nos permiten realizar una correcta decodificación. También se mostró que una manera más clara visualizar los diferentes códigos que se pueden definir para un alfabeto dada es por medio de un árbol. Los árboles nos permiten encontrar propiedades de los códigos de una manera geométrica.

El punto clave de una codificación óptima consiste en asignar las palabras más largas a los símbolos que son menos frecuentes en la cadena de caracteres, mientras que se deben asignar las palabras más cortas a aquellos símbolos cuya frecuencia es mayor en la cadena de caracteres. Una manera de realizar dicha asignación queda establecida por el algoritmo de Huffman. El algoritmo se basa en ir construyendo códigos bajo los cuales se puede asegurar que son óptimos paso a paso.

La codificación de una cadena de símbolos del longitud n tiene en promedio una longitud de nh, donde h es la entropía asociada a la transformación corrimiento σ con respecto a la medida de Bernoulli. La prueba resulta ser básicamente una consecuencia de la AEP. Para probar de manera formal la AEP, se necesitaron algunos resultados y conceptos de la Teoría Ergódica.

En primer lugar se definió la noción de medida T-invariante, o bien, la noción de transformación que preserva la medida. Se probó que para espacios métricos compactos y transformaciones continuas, siempre existen las medidas T-invariantes.

Posteriormente se probaron los resultados principales de la Teoría Ergódica, a saber, el Teorema de la Recurrencia de Poincaré y el Teorema Ergódico de Birkhoff. Ambos teoremas nos dan información acerca de la frecuencia en que la órbita de un punto visita un conjunto medible dado. El primero sólo afirma que la órbita de un punto regresa un número infinito de veces a un conjunto medible, mientras que el segundo nos permite calcular de manera explícita la frecuencia con que lo hace.

La propiedad más importante que pueden poseer las transformaciones es la ergodicidad. La ergodicidad describe un comportamiento de los puntos en el espacio bajo la acción de una transformación. Las órbitas de los puntos de una transformación ergódica recorren todos los conjuntos medibles del espacio. Dicho comportamiento trae importantes consecuencias sobre el Teorema Ergódico de Birkhoff pues permite calcular el límite de las frecuencias relativas de manera explicita.

Finalmente, para probar la AEP, se ocupó del Teorema de Shannon-McMillan-Breiman, que es el teorema fundamental de la Teoría de la Información. Para entender el Teorema de Shannon-McMillan-Breiman se requiere de los conceptos de entropía de una partición medible y de entropía de una transformación medible. La prueba ocupa el Teorema Ergódico de Birkhoff. En el Teorema de Shannon-McMillan-Breiman cobra de nueva cuenta suma importancia la hipótesis de la ergodicidad. Aunque se puede probar un resultado parcial para calcular la entropía de una transformación, sólo cuando se supone la ergodicidad se tiene un resultado directo sobre la entropía de una transformación o, como también le llamamos en algún momento, una fuente. La prueba de la AEP supone la ergodicidad de la transformación.

Bibliografía

[1] Arnold V. I., Avez A. c1989, "Ergodic Problems of Classical Mechanics",Redwood city, california : Addison-wesley, c1989.

[2] Barreira, Luis, 2012, "Ergodic theory, hyperbolic dynamics and dimension theory ",Heidelberg : Springer Verlag. 2.2

[3] Bartle, Robert G., 1995, "The elements of integration and lebesgue measure", New York : J. Wiley. (document)

[4] Brin, Michael, y Stuck, Garret 2002, "Introduction to Dynamical Systems", Cambridge : Cambridge University Press.

[5] Bavaud, Chappelier y Kohlas 2006, "An Introduction to Information Theory and Applications", URL=http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/diuf.unifr.ch/tcs/courses/it04-05/script/information-theory.pdf.

[6] Choe, Geon Ho, 2005, "Computational ergodic theory", Berlin : Springer Verlag.

[7] Collet, Pierre, y Eckmann, Jean-Pierre 2006, "Concepts and Results in Chaotic Dynamics. A Short Course", Berlin : Springer Verlag.

[8] Cover, Thomas M. y Thomas, Joy, "Elements of Information Theory", Hoboken, New Jersey : Wiley-Interscience. 5.3

[9] Dajani, Karma y Dirksin, Sjoerd 2008, "A Simple Introduction to Ergodic Theory", *Lecture Notes. Masters in Stochastics and Financial Mathematics. University of Utrecht. Ergodic theory (2012/2013)*, URL=http://www.staff.science.uu.nl/~kraai101/lecturenotes2009.pdf. 2.2

[10] Dietel, Harvey y Dietel, Paul 1995, "Como programar en C/C++", trad. Sánchez García Gabriel. México : Pearson/Educación, 1995.

[11] Grabinsky, Guillermo, 2009, "Teoría de la medida", México D.F.: UNAM, Facultad de Ciencias. (document)

[12] Halmos, Paul R., 1956, "Lectures on Ergodic Theory", New York: American Mathematical Society. (document)

[13] Halmos, Paul R., 1974, "Measure Theory", New York : Springer Verlag.

[14] Khinchin, A. I. 1957, "Mathematical Foundations of Information Theory", New York: Dover.

[15] Lind D. Brian M. 1995, "An Introduction to Symbolic Dynamics and Coding", Cambridge : Cambridge University Press.

[16] Markarian, Roberto. 2002, "Billares. Modelos con Dinámicas Caóticas" *Boletín de la Asociación Matemática Venezolana*, Vol.10, No. 2.

[17] Pérez, Francisco. 2010, "Algunos aspectos probabilísticos de teoría ergódica". Tesis. UNAM, Facultad de Ciencias, institución que otorga el título.

[18] Petersen, Karl. 1983, "Ergodic theory", Cambridge : Cambridge University Press.

[19] Rincon, L. 2004, "¿Qué es la esperanza condicional?" *Miscelánea Matemática*, 37:17-30. 2.2.3

[20] Rincon, L. 2007, "Curso Intermedio de Probabilidad", México D.F.: UNAM, Facultad de Ciencias.

[21] Salomon, David 2004, "Data Compression The Complete Reference", New York : Springer Verlag. (document)

[22] Ugalde, E. 2007, "De la mecánica estadística a la teoría ergódica" *Revista mexicana de física*, 53:191–194. 3.1

[23] Walters, Peter, 1982, ""An Introduction to Ergodic Theory", New York : Springer Verlag.

[24] Willard, Sthepen, 1941, "General Topology", Mineola, New York : Dover.

Índice alfabético